

(19) 世界知的所有権機関  
国際事務局



(43) 国際公開日  
2005 年10 月6 日 (06.10.2005)

PCT

(10) 国際公開番号  
WO 2005/093631 A1

(51) 国際特許分類<sup>7</sup>: G06F 19/00, C12M 1/00,  
C12N 15/09, C12Q 1/68, G06F 17/30  
(21) 国際出願番号: PCT/JP2005/005290  
(22) 国際出願日: 2005 年3 月23 日 (23.03.2005)  
(25) 国際出願の言語: 日本語  
(26) 国際公開の言語: 日本語  
(30) 優先権データ:  
特願2004-093301 2004 年3 月26 日 (26.03.2004) JP  
(71) 出願人 (米国を除く全ての指定国について): 株式会社  
バイオシンクタンク (BIO-THINK TANK CO., LTD.)  
[JP/JP]; 〒1130033 東京都文京区本郷三丁目 3 2 番  
6-7 0 3 号 Tokyo (JP).

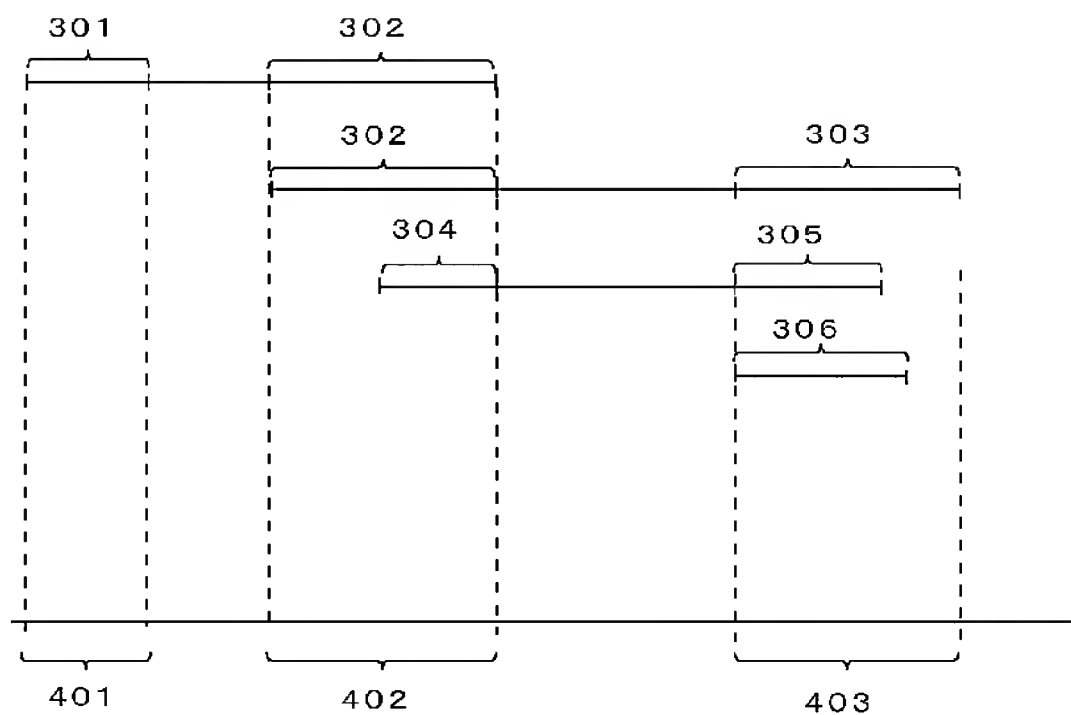
(72) 発明者; および  
(75) 発明者/出願人 (米国についてのみ): 森下真一 (MOR-  
ISHITA, Shinichi) [JP/JP]; 〒1790073 東京都練馬区田  
柄 1 丁目 1 9-7 Tokyo (JP). 山田智之 (YAMADA, To-  
moyuki) [JP/JP]; 〒1130033 東京都文京区本郷 5 丁目  
1 4-8 テラス本郷 3 0 2 号室 Tokyo (JP). 内藤雄樹  
(NAITO, Yuki) [JP/JP]; 〒1020082 東京都千代田区一  
番町 6-4-1 0 2 Tokyo (JP).  
(74) 代理人: 工藤一郎 (KUDO, Ichiro); 〒1000006 東京都  
千代田区有楽町 1-7-1 有楽町電気ビル (南館)  
9 階 Tokyo (JP).  
(81) 指定国 (表示のない限り、全ての種類の国内保護が  
可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR,  
BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM,

[続葉有]

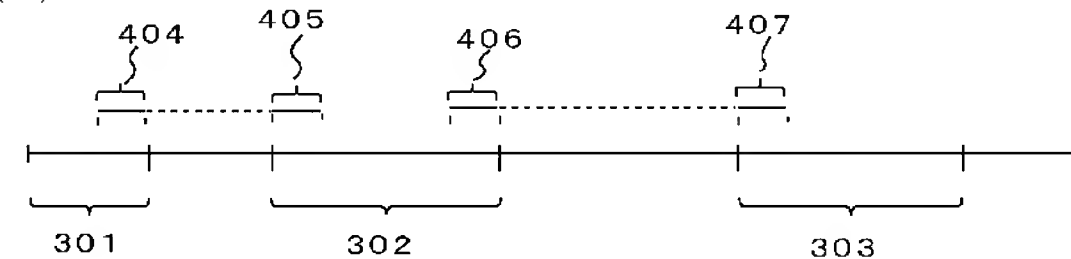
(54) Title: METHOD OF SEARCHING SPECIFIC BASE SEQUENCE

(54) 発明の名称: 特異的塩基配列探索方法

(A)



(B)



(57) Abstract: It is intended to efficiently determine a base sequence specifically appearing in an expression gene. For this, providing that the expression gene consists of exons (301)...(306) and especially that exon (301) is united with exon (302) and exon (302) with exon (303), an aggregate of base sequences (401) (403) being a union of exon base sequences (301)...(305) and a boundary base sequence obtained by uniting together base sequences(404) and (405) and base sequences (406) and (407) respectively existing over boundaries between exon (301) and exon (302) and between exon (302) and exon (303) is formed, and the aggregate is searched. If a base sequence is one specifically appearing in the expression gene, the number of search results is 1 and otherwise, the number is plural.

(57) 要約: 発現遺伝子に特異的に現れる塩基配列を効率よく決定する。かかる目的を達成するために、発現遺伝子が、エクソン (301) … (306) で構成され、特に、エクソン (301) とエクソン (302)、エクソン (302) とエクソン (303) が接合すると

[続葉有]

WO 2005/093631 A1



DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告書

(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE,

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

した場合、エクソンの塩基配列 (301) … (305) の和集合である塩基配列 (401) … (403) と、エクソン (301) とエクソン (302)、エクソン (302) とエクソン (303) の境界にまたがって存在する塩基配列 (404) と (405)、(406) と (407)、を接合して得られる境界塩基配列と、の集合を作り、この集合に対して検索を行なう。もし、発現遺伝子に特異的に現れる塩基配列であれば、検索結果数は1となり、そうでなければ、複数となる。

## 明 細 書

## 特異的塩基配列探索方法

## 技術分野

[0001] 本発明は、遺伝子の塩基配列に特異的に現れる塩基配列を検索する方法、装置、プログラムなどに関する。

## 背景技術

[0002] ワトソンとクリックとによるDNA (Deoxyribo Nucleic Acid) の構造の解明に基づき、塩基配列に基づく遺伝子情報の研究が発展している。DNAは、アデニン(A)、シトシン(C)、グアニン(G)、チミン(T)の塩基のいずれかを含むヌクレオチドが並んでいる構造を持ち、細胞の核の中では、通常、AとT、GとC、の結合により、二重らせんの構造となっている。遺伝子を表現するDNAのヌクレオチドの配列(以下、「遺伝子配列」と呼ぶ)が、RNA (Ribonucleic Acid) に転写され、スプライシングを経て、mRNA (messenger RNA) が生成され、たんぱく質の合成がされることが知られている。RNAは、D-リボースを糖成分として、アデニン(A)、シトシン(C)、グアニン(G)、ウラシル(U)を塩基とする核酸である。遺伝子配列のうち、たんぱく質の情報を持つ部分がエクソンと呼ばれ、そうでない部分は、イントロンと呼ばれる。したがって、スプライシングにより、RNAのイントロン部分が切除されることとなる。

[0003] 近年、RNA干渉と呼ばれる現象が発生することが知られるようになった。RNA干渉とは、細胞内の2本鎖RNAの存在により、特定の配列のmRNAを破壊し、遺伝子の発現を抑制する現象である。この現象は、最初、線虫の細胞を用いた実験で見られた。その後、この現象は、哺乳動物細胞でも起きることが知られるようになり、注目を集めることとなった。人為的にRNA干渉を起こすことにより、特定の遺伝子の働きを抑制することにより、その特定の遺伝子の働きを調べることができるからである。また、RNA干渉の発見により、特定の遺伝子の働きを抑制する効果を発揮する薬を開発できる可能性も生まれてきた。

[0004] 図1は、RNA干渉の過程の概略を示す図である。RNA干渉は、以下のようなプロセスを経て発生すると考えられている。およそ21から23塩基対の長さのsiRNA(

short interfering RNA) 101がマルチ・タンパク質複合体と結合し、RISC (RNA-induced silencing complex) 102を形成する。RISCは、そのsiRNAと相同性を持つmRNA 103と結合し、そのmRNAを分解することにより、そのmRNAが機能しなくなる(図1において、断片104、105が、mRNAが分解されてできた断片を表わしている)。ここで、「二つの塩基配列が相同性を有する」とは、2つの塩基配列が相補性を有しているか、または、不完全な相補性を有していることをいう。ここに「相補性」とは、二つの塩基配列の全体において、AとT、GとC、AとUとの対が完全に形成されていることをいう。したがって、相同性とは、二つの塩基配列の一部に、AとT、GとC、AとUという三種類の相補性を有する塩基対以外の対が発生していることを意味する。なお、どのような場合に、二つの塩基配列の間に相補性を有する塩基対がどれだけの存在すれば、その二つの塩基配列が相同性を有すると判断されるかについて説明すると次のようになる。すなわち、RNA干渉の場合には、80%以上、好ましくは90%以上、さらに好ましくは95%以上の場合に、相同性を有すると判断される場合が多い。また、相補性を有する塩基対の割合のみならず、相補性を有する塩基列が塩基配列中にどれだけの個数連続して現れているかを考慮にいて、二つの塩基配列の間の相同性の有無を判断することもある。また、AとT、GとC、AとUとの3種類の相補性を有する塩基対に、GとUとの対が形成される可能性もあることが知られているので、GとUとの塩基対の存在も考慮に入れて相同性の有無を判断することもある。

[0005] したがって、RNA干渉を発生させ、目的とする遺伝子の働きを抑制するためには、siRNAの配列を決定することが重要である。すなわち、目的とする遺伝子だけに現れ、他の遺伝子の塩基配列と相同性を持たない、siRNAの配列を決定することが重要である。

[0006] なお、哺乳類においては、ある遺伝子の特定領域と相同性を有するsiRNAの全てがRNA干渉を起こすわけではないことが知られている。そのため、RNA干渉を発生させるためのsiRNAの塩基配列の評価方法が提案されている(例えば、非特許文献1参照。)。この知見からすると、本発明は、塩基配列の評価の前段階として実施されるべきものである。あるいは、塩基配列の評価を行なった後に、高い評価値が得られ

た塩基配列の中から本発明を実施して特定領域と相同性を有する塩基配列を得るようにしてもよい。

[0007] また、近年、マイクロアレイを用いた遺伝子解析や遺伝子診断などが実施されている。「マイクロアレイ」とは、長さが15から30塩基程度のオリゴDNAをガラスなどの基板上に合成したDNAチップの一種である(例えば、非特許文献2参照。)

[0008] 図2は、マイクロアレイを用いた遺伝子解析や遺伝子診断などの過程を例示する。ガラスなどの基板上に合成したオリゴDNAを持つマイクロアレイ201上に、蛍光色素などの標識203を付加されたDNA(202)を流すと、そのDNAと相補性あるいは相同性を持つマイクロアレイ上のオリゴDNAとが結合(ハイブリダイズ)する(符号204)。どの場所のオリゴDNAとハイブリダイズしたかを、標識の蛍光色素による蛍光を検出することにより、DNA(202)の種類などを判定する。図2では、マイクロアレイ上に数本のオリゴDNAしか示されていないが、実際のマイクロアレイは、縦横の長さが0.5インチ程度の領域に万のオーダーでオリゴDNAが配置される。

[0009] したがって、どのような塩基配列を持つオリゴDNAをマイクロアレイに配置するかを決めることは、マイクロアレイの設計において、極めて重要な工程である。

非特許文献1:Angela Reynolds他著、“Rational siRNA design for RNA interference”、Nature Biotechnology、Published online 1 February 2004.

非特許文献2:杉本直己著、“遺伝子化学”、19ページ、株式会社化学同人発行、2002年

## 発明の開示

### 発明が解決しようとする課題

[0010] 本発明が解決しようとする課題は、与えられた遺伝子に特異的に現れる塩基配列を効率よく決定することである。「特異的」とは、その遺伝子にだけ現れ、他の遺伝子には現れないことを意味する。これにより、与えられた遺伝子だけを抑制するためのsiRNAの塩基配列が得られる。また、与えられた遺伝子だけを検出するオリゴDNAの配列が得られる。

[0011] 既に遺伝子の塩基配列のデータベースが構築されているが、そのような既存のデ



データベースを使用して特異的に現れる塩基配列を決定するには困難が伴う。これについて以下説明する。

[0012] 図3は、DNA配列と、mRNAに転写される発現遺伝子配列と、の関係を示す。図3(A)には、4本のDNA配列の部分が示されているが、これらは、分かりやすさのために、一つのDNA配列全体の一部を示しており、同じ部分の塩基配列が上下に対応するように示されている。DNA配列には、発現遺伝子を構成するエクソンの部分と、発現遺伝子を構成しないイントロンの部分と、があることが知られている。図3(A)において、符号301、302、303、304、305、306の部分がエクソンであり、他の部分がイントロンであるとする。図3(B)は発現遺伝子配列を示す。図3(B)に示すように、一つのエクソンは、一つの発現遺伝子配列にだけ現れるとは限らず、複数の発現遺伝子配列に表れる場合がある。例えば、エクソン302は、エクソン301と接合されることにより、ある発現遺伝子を構成するが、エクソン303と接合されることにより、別の発現遺伝子を構成する。

[0013] また、エクソンの一部がエクソンとなっている場合がある。例えば、図3(A)においてエクソン302の一部が、エクソン304となり、また、エクソン303の一部が、エクソン305、エクソン306になっている。

[0014] したがって、発現遺伝子配列を格納するデータベースにおいては、一つのエクソン、またはその一部、の塩基配列が、複数の発現遺伝子配列に表れることになる。このため、例えば、エクソン302に特異的に現れる塩基配列を検索すると、検索の結果は一つではなく、複数あることになり、特異的に現れる塩基配列でないと判断されてしまう可能性がある。その可能性を排除するため、検索の結果が複数得られた場合には、検索の結果を精査して、特定のエクソンだけに特異的に現れる配列かどうかのチェックを別途行なう必要がある。

[0015] このような現象を避ける一つの方法としては、ゲノム配列全体に対して検索を行なうものがある。しかし、このような検索を行なうと、発現遺伝子配列のエクソンの境界をまたぐ塩基配列が検索されないことになってしまう。すなわち、発現遺伝子配列が、ゲノム配列中の複数のエクソンが接合してできる場合において、塩基配列の一部分が一つのエクソンに含まれ、その塩基配列の残りの部分が他のエクソンに含まれる場合、す

なわち、エクソンの端に位置する塩基であるエクソンの境界が、その塩基配列の中に含まれる場合には、その塩基配列はゲノム配列中にそのまま現れることはないので、検索されない。このため、ある塩基配列が、発現遺伝子配列のエクソンの境界をまたぐように複数回現れる場合には、その塩基配列が特異的なものでないという判断を行なうことができない。もしくは、エクソンの境界をまたぐような配列が特異的であったとしても、その配列が特異的であるという判断を行なうこともできない。

[0016]   そこで、本発明は、発現遺伝子に特異的に現れる塩基配列(より正確に言えば、一つのエクソンに特異的に現れる塩基配列、又は、エクソンが結合することにより発現遺伝子に特異的に現れる塩基配列)を効率よく検出する方法、装置、データベース、プログラムなどを提供することを目的とする。

#### 課題を解決するための手段

[0017]   本発明においては、塩基配列のデータベースとして、エクソンの塩基配列の和集合と、発現遺伝子のエクソンの境界をまたぐ塩基配列の集合と、の和集合を用いて検索を行なう。これにより、発現遺伝子配列に特異的に現れる塩基配列であれば、検索の結果は一となる。また、特異的に現れる塩基配列でなければ、検索の結果は複数となる。結果として、検索結果を調べるだけで、特異的に現れる塩基配列かどうかを直ちに判定することができる。これにより、課題が解決される。

[0018]   なお、発現遺伝子のエクソンの境界をまたぐ塩基配列は、適宜統合することが可能である。これにより、データベースのレコード数を減少させることもできる。

[0019]   また、相同性の程度を指定するために、検索の際に、いくつまでの塩基のミスマッチ(不適合)を許すかどうかを指定するようにしてもよい。また、加えて、相同性の程度を指定するために、不適合とみなす塩基の対を指定してもよい。また、不適合の発生の分布を指定してもよい。このように指定される分布の例としては、不適合でない塩基が連続する長さ(すなわち、塩基の対が連続して発生する長さ)がある。この長さがある程度以上の長さになると、RNA干渉においては、不適合の塩基配列があるにもかかわらずsiRNAがmRNAに結合してしまうと考えられている。そこで、そのような結合を排除するために、不適合でない塩基が連続する長さを指定する。

[0020]   なお、本発明においては、ゲノム配列中のどの部分がエクソンであり、また、イントロ

ンであるかについてが、検索に用いられる塩基配列のデータベースの構成に大きな影響を与える。以下の説明では、すでに研究された結果を用いることを主に想定しているが、今後の研究成果を取り入れて、塩基配列のデータベースを構成することが可能である。

### 発明の効果

- [0021] 本発明においては、エクソンの塩基配列と、エクソンの境界に現れる塩基配列と、から塩基配列集合を生成して、検索を行なうことにより、発現遺伝子に特異的に現れる塩基配列かどうかを検索結果数に基づいて決定できる。

### 発明を実施するための最良の形態

- [0022] 以下、本発明を実施するための最良の形態を、図を用いて、実施形態として説明する。なお、本発明は、これら実施形態に何ら限定されるものではなく、その要旨を逸脱しない範囲において、種々なる態様で実施し得る。

- [0023] (発明の概要)

実施形態について説明する前に、本発明の概要をいくつかの節に分けて説明する。

- [0024] 図4は、エクソンの和集合と、発現遺伝子のエクソンの境界をまたぐ塩基配列と、を説明するための図である。なお、以下では、発現遺伝子を構成するエクソンの境界をまたぐ塩基配列を、「境界塩基配列」ということにする。

- [0025] <第一節:エクソンの塩基配列の和集合>

図4(A)は、エクソンの塩基配列の和集合を説明するための図である。図4(A)には、図3(A)のように、4本のDNA配列の部分が示されているが、これらは、一つのDNA配列全体の一部を示しており、同じ部分の塩基配列が、上下に対応するように示されている。エクソン301、302、303、304、305、306が図のような関係にあるとする。すなわち、エクソン301と重なり、又は、包含関係にあるエクソンは他になく、エクソン302の一部として、エクソン304があり、エクソン303の一部にエクソン305、306があるとする。このような場合に、これらのエクソンの和集合として、配列401、402、403が得られる。すなわち、配列401は、エクソン301そのものであり、配列402は、エクソン302とエクソン304との和である。このエクソン304は、エクソン302の一部であ



るので、配列402は、エクソン302そのものとなる。同様に、配列403は、エクソン303そのものとなる。図4においては、エクソン302とエクソン304との関係のように、あるエクソンが他のエクソンを包含している関係にある場合が示されている。他の関係としては、包含ではなく、二つのエクソンの塩基配列の一部だけ重なっている場合がある。この場合については、後に図6、図7などを用いて説明する。

[0026] <第二節:境界塩基配列>

図4の下部は、境界塩基配列を説明するための図である。エクソン301とエクソン302とが接合して発現遺伝子を構成する場合、その接合における境界の左右の部分404と405とを接合した塩基配列が、境界塩基配列となる。同様に、エクソン302とエクソン303とが接合する場合、部分406と407を接合した塩基配列が、境界塩基配列となる。なお、ここでの境界塩基配列の長さは、発現遺伝子配列に特異的に現れるかどうかを調べるための検索が行なわれる塩基配列の長さである。その長さをNとすると、境界塩基配列は、N-1通りあることになる。

[0027] 図5は、N-1通りの境界塩基配列を例示する。エクソン501とエクソン502とが接合して発現遺伝子を構成するとする。この場合、エクソン501の右端のN-1mer(「mer」は、塩基配列の長さの単位であり、1塩基の長さを1merとする)の部分503と、エクソン502の左端の1merの部分504と、を接合することにより、境界塩基配列が一つ得られる。以下、同様に、N-2merの部分505と2merの部分506、…、2merの部分507とN-2merの部分508、1merの部分509とN-1merの部分510、のN-2通りの塩基配列が得られる。これらN-1通りの塩基配列は、包含関係ではなく一部だけが重なっている関係にあり、エクソンの和集合を求める場合のように、統合して一つにまとめることが可能である。

[0028] <第三節:塩基配列の統合>

図6は、塩基配列の統合を説明するための図である。すなわち、塩基配列601と塩基配列602とが、部分603の重なっている関係にある場合、塩基配列601と塩基配列602とを統合して、塩基配列604が得られることが示されている。塩基配列604は、塩基配列601から部分603を除いた部分、部分603、塩基配列602から部分603を除いた部分の3つを接合することにより得られる。

[0029] < 第四節：塩基配列の統合の処理 >

図7は、統合を正確に説明するための図である。図7の上部に示すように、DNAの塩基配列を構成する塩基は、DNAの端（例えば、DNAの化学構造により、「5′末端」と呼ばれる端）の塩基を1として順に番号を付けることができる。例えば、端点701を5′末端とし、端点702を3′末端として、端点701の塩基より、1、2、3、…のように塩基に番号をつけることができる。このような番号を、塩基位置ということにする。図7の下部703において、例えば、塩基配列704に現れる塩基Aの上に1024が付されているのは、その塩基Aは、DNAの5′末端より、1024番目に現れることを示す。塩基配列704と705とが、一部だけが重なっている関係にあるとする。すなわち、塩基配列1026番目と1027番目との部分重なっている。この場合、塩基配列704と705とを統合することにより、塩基配列706が得られる。

[0030] 図8は、塩基配列の和集合、特に統合、を計算するために用いるテーブルを例示する。ここでいう「計算」は計算機を用いてプログラムを動作させて行なうのが好適である。その場合には、テーブルとしては、データベース管理システムなどで管理されるようになっていてもよい。図8のテーブルは、「左端位置」と「右端位置」という名の列を有している。各行は、エクソンの塩基配列の左端と右端の塩基位置を格納する。また、エクソンの境界をまたぐ塩基配列の左端と右端の塩基位置を格納してもよい（後に説明するように、エクソンの境界をまたぐ塩基配列の統合には、やや複雑な操作が必要となる場合がある。図8のテーブルを使用することができるのは、限られた場合である）。なお、テーブルの各行には、行801には1が、行802には2が、という具合に行番号が付いているとする。従って、行801を「第1行目」、行802を「第2行目」という。

[0031] また、図8に例示されたテーブルに格納される各行に関連づけて、エクソンの属性情報が蓄積されていてもよい。例えば、図8のテーブルの行番号と、エクソンの属性情報と、を関連付けて蓄積する別のテーブルがあってもよい。あるいは、エクソンの属性情報は、図8に例示されたテーブルに列を追加してその列に格納されてもよい。ここに、「属性情報」とは、(1)エクソンの配列位置を示す情報または(2)エクソンが構成する遺伝子を識別する情報を、含む情報をいう。「エクソンの配列位置を示す情報」とは、エクソンがゲノム配列のどの位置に存在するかを示す情報である。例えば、D

NAの端からの位置である。この情報は、図8に例示されたテーブルの左端位置または右端位置の列に格納されているが、和集合を求める際に、左端位置または右端位置の列に格納されている値が変化するので、別に格納してもよい。また、「エクソンが構成する遺伝子を識別する情報」とは、そのエクソンの塩基配列を含む遺伝子を表わす情報、例えば、遺伝子の名前など、である。また、エクソンの配列位置を示す情報、エクソンが構成する遺伝子を識別する情報、以外には、エクソンの長さなどがある。

- [0032] 図9は、塩基配列の和集合、特に統合、を計算するための処理のフローチャートを例示する。上述したように、「計算」は計算機を用いて行なうのが好適である。したがって、図9に例示されるフローチャートの処理は、計算機で行なうのが好適である。ステップS901において、左端位置という名前の列の値により、昇順に行をソートする。すなわち、図8に例示された表の行を、 $N+1$ 行目の左端位置という名前の列の値が $N$ 行目の左端位置という名前の列の値より小さくならないように、並び替えることを行なう。次にステップS902において、変数 $r$ に2を代入する。変数 $r$ は、現在、何行目の処理を行なっているかを示す変数である。
- [0033] ステップS903において、 $r$ の値が、全行数の値以下であるかどうかを判断する。すなわち、第 $r$ 行目がテーブルに存在するかどうかを判断する。もし、そうならば(ステップS903:Yへ分岐する場合)、ステップS904以下を行なう。そうでなければ(ステップS903:Nへ分岐する場合)、全ての行に対する処理が終わったことになる。
- [0034] ステップS904において、第 $r$ 行目が表わす塩基配列と第 $(r-1)$ 行目が表わす塩基配列が包含関係または一部が重なる関係にあるかどうかを調べる。すなわち、第 $(r-1)$ 行目の左端位置の列の値 $\leq$ 第 $r$ 行目の左端の列の値、かつ、第 $r$ 行目の左端位置の列の値 $\leq$ 第 $(r-1)$ 行目の右端位置の列の値、が成立するかどうかを調べる。ステップS905において、成立する場合(ステップS905:Yへ分岐する場合)には、ステップS906へ分岐し、そうでなければ(ステップS905:Nへ分岐する場合)、ステップS909へ分岐する。
- [0035] ステップS906において、第 $r$ 行目の左端位置の列へ、第 $(r-1)$ 行目の左端位置の列の値を代入する。ステップS907において、第 $r$ 行目の右端位置の列の値が第 $(r-$

1) 行目の右端位置の値より小ならば、第 $r$ 行目の右端位置の列へ、第 $(r-1)$ 行目の右端位置の値を代入する。ステップS906とステップS907により、第 $(r-1)$ 行目と第 $r$ 行目とが表わす塩基配列を統合したものが、第 $r$ 行目により表わされるようになる。したがって、第 $(r-1)$ 行目は不要となるので、ステップS908により、第 $(r-1)$ 行目を削除する。これにより、全行数の値は1減ることとなる。その後、ステップS903へ戻る。なお、ステップS908において、第 $(r-1)$ 行目を削除して消滅させるのではなく、第 $(r-1)$ 行目を別のテーブルに移動して蓄積してもよい。これにより、例えば、エクソンの位置が、元来どの配列に由来しているものであるかという情報をその別のテーブルに蓄積することができ、検索が可能となる。

[0036] なお、ステップS907において、第 $r$ 行目に関連付けて蓄積されている属性情報を、第 $(r-1)$ 行目に関連付けて蓄積されている属性情報にマージすることを行なってもよい。マージの例としては、第 $r$ 行目に関連付けて蓄積されている属性情報を表現する文字列と、第 $(r-1)$ 行目に関連付けて蓄積されている属性情報を表現する文字列と、を接続する。このように接続して得られた文字列を第 $(r-1)$ 行目に関連づけて蓄積される属性情報としてもよい。例えば、第 $(r-1)$ 行目に関連付けて、「A、B」のように区切りとして「、」が用いられて「A」と「B」とが蓄積され、第 $r$ 行目に関連付けて「C」が蓄積されていれば、「A、B」と「C」とを、区切りを示す「、」とともに接続して得られる「A、B、C」を第 $(r-1)$ 行目に関連付けて蓄積してもよい。このようにすることにより、エクソンの和集合の要素がどのエクソンに由来しているか、例えば、どの遺伝子に関係しているか、を容易に知ることができる。

[0037] ステップS909においては、次の行に対する処理を行なうために、 $r$ の値を1増加させ、ステップS903へ戻る。

[0038] <第五節：境界塩基配列の統合が直ちに求められる場合>

図10は、二つのエクソンが接合して発現遺伝子を構成する場合における $N-1$ 通りの境界塩基配列を統合した塩基配列の求め方を例示する。エクソン1001とエクソン1002とが接合して発現遺伝子を構成するとする。この場合、エクソン1001とエクソン1002との境界における境界塩基配列を統合した塩基配列は、エクソン1001の右端の $N-1$ merの塩基配列1003と、エクソン1002の左端の $N-1$ merの塩基配列1004



を接合した $2N-2$ merの塩基配列となる。ただし、図10においては、エクソン1001とエクソン1002のそれぞれの長さが $N-1$ mer以上である必要がある。

[0039] <第六節:境界塩基配列の統合が直ちに求められない場合>

図11は、長さが $N-1$ mer未満のエクソンが存在する場合を例示する。図11において、符号1101、1102、1103、1104を付した部分がエクソンであるとし、エクソン1101、1102、1103が接合して一つの発現遺伝子を構成し、エクソン1101、1102、1104が接合して別の発現遺伝子を構成するとする。また、エクソン1102の長さは $N-1$ mer未満とし、エクソン1103とエクソン1104は、一部が重なった関係にあるとする。符号1105、1106、1107、1108を付した部分はイントロンであるとする。

[0040] この場合、境界塩基配列を求めると、符号1109、1110が付されたものの実線部分に相当するものが得られる。発現遺伝子に特異的に現れる塩基配列かどうかを判断するための検索は、エクソン1101、1102、1103、1104の和集合に、これらの境界塩基配列の集合を和として加えた集合に対して行なうことになる。あるいは、これらの境界塩基配列の集合の代わりに、境界塩基配列の集合に対して次のような統合の操作を行なって得られる塩基配列の集合を用いてもよい。

[0041] <第七節:境界塩基配列の統合を求める一般的な処理>

図12は、統合の操作を行なうために使用するテーブルを例示する。テーブルは、「発現遺伝子」、「左端位置」、「右端位置」の列からなっている。「発現遺伝子」の列は、境界塩基配列が現れる発現遺伝子を識別する識別子を格納する。図12では、発現遺伝子を構成するエクソンの符号を並べたものにより、そのような識別子が表わされている。「左端位置」と「右端位置」とは、図8のテーブルにおける意味と同じ意味を持ち、境界塩基配列の左端の塩基の位置と、右端の塩基の位置と、を格納する。なお、統合の操作も、計算機でプログラムを動作させることにより実行することが可能である。その場合、テーブルは、データベース管理システムにより管理されて操作が行なわれるようになっていてもよい。また、そのようなプログラムはフレキシブルディスク、光ディスク、メモリスティックなどの媒体に記録することもできる。

[0042] まず、一つの境界塩基配列に対応して、図12のテーブルの行が一つ作られるが、境界塩基配列の集合がテーブルに格納されるようにするために、「左端位置」と「右



端位置」の列の値の組がユニークになるようにする。すなわち、「左端位置」と「右端位置」の列の値の組が複数回テーブルに現れないようにする処理を行なう。この処理を高速に行なうためには、例えば、左端位置の列と右端位置の列との組みに対して索引を定義しておき、テーブルに新たな行を追加しようとする際には、その索引を参照して、すでにテーブルに格納されている行の左端位置と右端位置という列の値の組に同じものがあるかどうかを調べるようにすればよい。ここにいう索引は、キーとしてテーブルの左端位置という名前の列と右端位置という名前の列との組の値を持ち、バリューとしてテーブル行番号やテーブルの行を一意に特定する列の値を持つ。もし、索引を参照して、新たに追加しようとする行の左端位置と右端位置という名前の列の値の組が同じなるような行が既にテーブルにあれば、テーブルに行を追加することはしない。もし、新たに追加しようとする行の左端位置と右端位置という名前の列の値の組が同じなるような行がまだ格納されていなければ、テーブルに行を追加する。これにより、境界塩基配列の集合が得られる。

[0043] 次に、境界塩基配列の集合の要素の統合を行なう。この統合の際には、発現遺伝子の列の値が同じものの間で統合を行なう。すなわち、エクソン1101、1102、1103の境界塩基配列は、エクソン1101、1102、1103から構成される発現遺伝子の境界塩基配列と統合することとし、エクソン1101、1102、1104から構成される発現遺伝子とは統合しないようにする。このために、例えば、テーブルにおいて、発現遺伝子の列の値でソートを行ない、発現遺伝子の列の値が同じ行の集まりを作ることによりテーブルを分割して、それぞれの分割に対して、図9のフローチャートで示される処理を適用する。このように発現遺伝子の列の値が同じものの間で統合を行なうのは、発現遺伝子にあり得ない塩基配列が生成されることを防ぐためである。このような処理の結果、符合1113、1114が付された塩基配列が得られる。

[0044] 図13は、以上説明した境界塩基配列の集合に対する統合の処理のフローチャートを例示する。まず、最初のステップとして、左端位置、右端位置の列の値の組に重複が発生しないように、境界塩基配列の情報をテーブルに付加する。次のステップとして、発現遺伝子の列の値が同じである行の集合ごとに、統合の操作を行なう。すなわち、テーブルを、発現遺伝子の列の値が同じになるようにテーブルをグルーピングす

ることにより(例えば、SQL(Structured Query Language)におけるgroup by節を用いることにより)テーブルをいくつかの小テーブルに分割し、それぞれの小テーブルに対して図9のフローチャートで示される処理を適用する。

[0045] (実施形態1:主に請求項20、24に対応する)

図14は、本発明の実施形態1に係る塩基配列集合生成方法の処理のフローチャートを例示する。本実施形態に係る塩基配列集合生成方法は、候補塩基長取得ステップと、エクソン塩基配列集合取得ステップと、境界塩基配列集合生成ステップと、和集合生成ステップと、を含む。これらのステップは、図14に例示されたフローチャートのS1401、S1402、S1403、S1404にそれぞれ対応する。以下の説明から分かるように、これらのステップは、計算機にプログラムを動作させて実行させることが可能である。また、そのようなプログラムをフレキシブルディスク、光ディスク、メモリスティックなどの媒体に記録することも可能である。

[0046] 「候補塩基長取得ステップ」(S1401)は、発現遺伝子の塩基配列に特異的に現れる塩基配列の候補の塩基配列の長さ(以下、「候補塩基配列長」という。)を取得するステップである。取得される候補塩基配列長は、本実施形態に係る塩基配列集合生成方法により生成される塩基配列の集合が、siRNAの設計を目的とするならば、その上限は、好ましくは30以下、より好ましくは、22以下、さらに好ましくは20以下であり、その下限は、好ましくは13以上、より好ましくは16以上、さらに好ましくは18以上である。例えば、19が好適な値である。また、その塩基配列の集合が、マイクロアレイのオリゴDNAの設計を目的とするならば、その上限は30以下であるのが好ましい。

[0047] 「エクソン塩基配列集合取得ステップ」(S1402)は、エクソンの塩基配列の和集合を取得する。本明細書において「取得」という単語は、生成の意味を含むとする。もし、ここでエクソンの塩基配列の和集合を生成するのであれば、上記の第四節で述べたように生成する。

[0048] 「境界塩基配列集合生成ステップ」(S1403)は、境界塩基配列集合を生成する。「境界塩基配列集合」とは、複数のエクソンから構成される発現遺伝子におけるエクソンの境界にまたがって存在する塩基配列を示す情報であって、候補塩基配列長取得ステップで取得された長さと同じ長さの塩基配列を示す情報、からなる集合に対し

て、発現遺伝子が同じで、塩基配列の位置が重複する塩基配列を示す情報を統合することにより得られる集合である。具体的には、上記の第五節、又は、第六節、第七節で説明した処理により得られる塩基配列の集合である。

[0049] 「和集合生成ステップ」(S1404)は、エクソン塩基配列集合取得ステップで取得された塩基配列の集合と、境界塩基配列集合生成ステップで生成された塩基配列の集合と、の和集合を生成するステップである。このステップにおける和集合は、基本的には、単純な集合の和を取る操作で得られるものである。ただし、例外として、単純な集合の和の操作にならない場合が2つある。まず、エクソンの塩基配列の和集合の要素である塩基配列であって、発現遺伝子の端に配置され、 $N-1$ mer以下のものがある場合は、そのような塩基配列は境界塩基配列またはそれを統合した塩基配列に含まれている(すなわち、包含関係にある)ので、そのような塩基配列を除去する必要がある。また、エクソンの塩基配列の和集合の要素である塩基配列であって、発現遺伝子の端ではなく中間に配置され、 $2N-2$ mer以下のものがある場合には、そのような塩基配列が、境界塩基配列またはそれを統合した塩基配列に含まれる可能性がある( $N-1$ mer以下である場合には必ず含まれる)ので、そのような塩基配列が存在すれば除去する。

[0050] 図15は、図14の和集合生成ステップS1404で得られた塩基配列を格納したテーブルを例示する。例えば、「左端位置」の列には塩基配列の左端の塩基の塩基配列のDNA配列における位置を格納し、「塩基配列」の列には、塩基配列を格納する。他に、発現遺伝子の識別子などの情報を格納するための列があってもよい。

[0051] 本実施形態により生成される塩基配列の集合に対して検索を行なうことにより、与えられた遺伝子に特異的に現れる塩基配列を効率よく決定できることとなる。すなわち、特異的に現れる塩基配列であれば、その塩基配列を用いて塩基配列の集合を検索すると、検索結果は1となり、そうでなければ、検索結果は複数となる。

[0052] (実施形態2:主に請求項1、2、23に対応する)

図16は、本発明の実施形態2に係る特異的塩基配列探索方法のフローチャートを例示する。本実施形態に係る特異的塩基配列探索方法は、特異的塩基配列候補取得ステップと、塩基配列検索ステップと、判断ステップと、を含む。以下の説明から分

かるように、これらのステップは、プログラムによって計算機に実行させることが可能である。また、そのようなプログラムをフレキシブルディスク、光ディスク、メモリスティックなどの媒体に記録することも可能である。

[0053] 「特異的塩基配列候補取得ステップ」(S1601)は、特異的塩基配列候補を取得する。「特異的塩基配列候補」とは、発現遺伝子の塩基配列に特異的に現れる塩基配列の候補である。任意の塩基配列を候補としてもよいが、例えば、従来技術として知られる方法により塩基配列に対して特異的に現れる可能性が高いかどうかの評価を行ない、特異的に現れる可能性が高いという評価値が得られた塩基配列を候補としてもよい。ここにいう従来技術として知られている方法としては、(1)発現遺伝子の塩基配列情報と同一または類似の塩基配列を、NCBIのRefSeqなどのデータベースで公開されている塩基配列情報などから、例えば、BLAST、FASTA、ssearchなどの既存のホモロジ検索手段を用いて検索し、(2)検索された塩基配列のうち発現遺伝子とは無関係の遺伝子の塩基配列情報の総数や発現遺伝子とは無関係の遺伝子の塩基配列情報に付された同一類似の度合いを示す値—例えば、BALST、FASTA、ssearchの場合における「E value」—に基づいて、同一または類似の度合いを示す値の逆数の総和を算出して、その総和を求め、(3)その総和に基づいて—例えば、総和の大小などに基づいて—発現遺伝子の塩基配列に特異的であるか否かを判断する方法がある。特異的塩基配列候補取得ステップを計算機に実行させるためには、キーボードなどから入力された特異的塩基配列候補を表わす文字列などを読み取ることが計算機に行なわせる。

[0054] 「塩基配列検索ステップ」(S1602)は、塩基配列集合の中から、適合塩基配列を検索する。「塩基配列集合」とは、エクソンの塩基配列の和集合と、境界塩基配列の集合と、の和集合を含む集合である。塩基配列集合は、例えば、第一節で説明したエクソンの塩基配列の和集合と、第二節で説明した境界塩基配列の集合と、の和集合である。あるいは、実施形態1に係る塩基配列集合生成方法にて生成された集合であってもよい。エクソンの塩基配列の和集合については、エクソンの塩基配列に対して第四節で説明した統合の処理を行なって得られるものであってもよい。また、塩基配列集合は、ゲノム配列が解読されてない等の理由によって、エクソンであるか、



あるいは、その境界にまたがって存在するのかが不明な配列をさらに含んでもよい。場合によっては、塩基配列集合は、遺伝子の配列の集合全体となってもよい。また、第四節の終わりの部分で説明したように、エクソンの塩基配列の和集合の要素には、エクソンの配列位置を示す情報またはエクソンが構成する遺伝子を識別する情報が関連付けられていてもよい。

[0055] 「境界塩基配列」とは、第二節で述べた通りである。すなわち、複数のエクソンから構成される発現遺伝子におけるエクソンの境界にまたがって存在する塩基配列であり、特異的塩基配列候補の塩基配列と同じ長さの塩基配列である。「適合塩基配列」とは、特異的塩基配列候補取得ステップにて取得された特異的塩基配列候補が表わす塩基配列と適合する塩基配列である。ここに「2つの塩基配列が適合する」とは、2つの塩基配列を構成する塩基を、対ごとに比較を行なった結果、所定の二項関係を満たさない対が所定の数以下であることをいう。ここでいう二項関係とは、多くの場合、対を構成する塩基が合い等しいことをいう。すなわち、数学の集合論の言葉で説明すれば、二項関係が反射律のみを満たす場合である。また、塩基のGとUとが結合しやすいことを考慮に入れた二項関係を用いてもよい。また、二項関係のみに依存して適合かどうかを判断せず、適合する塩基の連続する数などを考慮に入れて、2つの塩基配列が適合するかどうかを判断してもよい。「所定の数以下」とは、例えば、20%以下、好ましくは10%以下、より好ましくは5%以下を意味してもよい。このような検索の方法については、バイオインフォマティクスの分野で研究が進んでおり、例えば、FASTA、BLAST、スミス-ウォーターマンダイナミックプログラミング法を使う方法など、計算機を用いて行なう方法が知られている(例えば、David W. Mount 著、“Bioinformatics: Sequence and Genome Analysis”, Cold Spring Harbor Laboratory Press, 2001年など参照)。

[0056] 「判断ステップ」(S1603)は、塩基配列検索ステップでの検索結果に、適合塩基配列が複数あるかどうかに基づいて、特異的塩基配列候補取得ステップにて取得された特異的塩基配列候補が特異的塩基配列であるか判断する。ここに「特異的塩基配列」とは、発現遺伝子に特異的に現れる塩基配列を意味する。判断ステップでは、検索結果の適合塩基配列が1であれば、特異的塩基配列候補が特異的塩基配列であ



ると判断すればよい。もし、検索結果の適合塩基配列が2以上であれば、特異的塩基配列でないと判断する。もし、検索結果の適合塩基配列が0であれば、類似のものが存在しないと判断する。検索結果の適合塩基配列が0となる場合には、このような特異的塩基配列候補は、何の効果ももたらさないと推定される。したがって、計算機に判断ステップを実行させるには、検索結果集合の数を取得させて判断をさせることになる。

[0057] (実施形態3:主に請求項3に対応する)

本発明の実施形態3は、実施形態2に係る特異的塩基配列探索方法において、境界塩基配列の集合を、第四節、第七節にあるように塩基配列の統合をして得られる集合としたものである。

[0058] すなわち、境界塩基配列の集合を、(1)複数のエクソンから構成される発現遺伝子におけるエクソンの境界にまたがって存在する塩基配列を示す情報であって、(2)特異的塩基配列候補の長さと同じ長さの塩基配列を示す情報、からなる集合に対して、発現遺伝子が同じで、塩基配列の位置が重複する塩基配列を示す情報を統合することにより得られる集合としたものである。なお、統合ができなくなるまで、すなわち、完全に統合の処理を行なう必要はない。また、統合の処理により、エクソンの塩基配列の和集合の中に、統合されて得られる塩基配列に含まれる塩基配列が現れる場合があり、そのような塩基配列を取り除く必要が出てくるのは、実施形態1で述べたとおりである。

[0059] 塩基配列を示す情報とは、例えば、図8に例示されたテーブルに格納された各行、あるいは、図12に例示されたテーブルに格納された各行、を意味する。

[0060] 本実施形態においては、統合が行なわれるので、検索が行なわれる要素を減少させることができ、集合のサイズを小さくすることができる。また、検索のスピードを向上させることができる。

[0061] (実施形態4:主に請求項4に対応する)

本発明の実施形態4は、実施形態2または3に係る特異的塩基配列探索方法に、適合許容数取得ステップを含ませた特異的塩基配列探索方法である。

[0062] 図17は、本実施形態に係る特異的塩基配列探索方法のフローチャートを例示する

。このフローチャートは、図16のフローチャートに適合許容数取得ステップであるS1702を追加したものである。

- [0063] 「適合許容数取得ステップ」とは、適合許容数を取得する。「適合許容数」とは、塩基配列集合の塩基配列と特異的塩基配列候補が表わす塩基配列との適合の度合いとして、いくつかの塩基の不適合まで許容するかを示す数値である。好ましくは、1、2、3、4、5のいずれかの値である。ここでいう「塩基の不適合」とは、塩基の対が所定の二項関係を満たさないことをいう。計算機に適合許容数取得ステップを実行させるためには、例えば、キーボードなどから入力されたり、画面に表示されたラジオボタンの選択により入力されたりする適合許容数を計算機に読み取らせる。
- [0064] 本実施形態においては、塩基配列検索ステップでは、適合許容数取得ステップで取得された適合許容数に基づいて検索が行なわれる。例えば、前に説明したBLASTなどを用いて検索を行なう。この場合、「適合許容数に基づいて」とは不適合となる塩基対の数が適合許容数以下になるように検索を行なうことを意味する。ただし、BLASTにおいては、通常、7塩基が連続して同じになる部分を用いて検索が行なわれるため、候補塩基配列長が19で適合許容数が3である場合には、図18の×の位置で塩基の不適合がある場合を検索することができない。そこで、特異的塩基配列候補において、×の位置の塩基を他の塩基に置き換えた塩基配列を生成して、特定記塩基配列候補が表わす塩基配列と一致しまたは相補性のある塩基配列を検索するようにしてもよい。なお、適合許容数を指定して検索を行なう方法としては、Tomoyuki YAMADA and Sinichi MORISHITA, "Computing Highly Specific and Noise-Tolerant Oligomers Efficiently, To appear in Journal of Bioinformatics and Computational Biology, Imperial College Pressに述べられている方法がある。
- [0065] (実施形態5:主に請求項5に対応する)
- 本発明の実施形態5として、塩基配列検索ステップにて不適合と判断する塩基の対を取得するステップを含む特異的塩基配列探索方法について説明する。
- [0066] 本実施形態に係る特異的塩基配列探索方法は、実施形態4に係る特異的塩基配列探索方法が、さらに、不適合塩基対取得ステップを含む方法である。
- [0067] 「不適合塩基対取得ステップ」とは、塩基配列検索ステップにおいて不適合と判断

する塩基の対を取得する。この取得は、計算機に接続されたキーボードから入力された塩基対を取得することにより行なわれたり、媒体に記録された塩基対を示す情報を読み込んだり、通信回線を通じて入力される情報を取得したりすることにより行なわれる。塩基配列検索ステップにおいては、同一の塩基でなければ不適合と扱うのが通常である。しかし、例えば、GとUとが結合して対を形成することが知られているので、GとUとの対を不適合とみなしたくない場合もある。そこで、本実施形態においては、不適合であると判断する塩基の対を取得することができるようにする。なお、不適合であると判断する塩基の対を取得するかわりに、適合すると判断する塩基の対を取得することにより、間接的に不適合であると判断する塩基の対を取得してもよい。また、取得される塩基の対は、適合あるいは不適合の程度を関連付けて取得されるようになってよい。例えば、同じ塩基の対であれば1という値を割り当て、例えば、GとUの対には、0.5という値を割り当ててもよい。なお、不適合塩基対取得ステップは、塩基配列検索ステップS1703が実行されるまでに実行される。例えば、不適合塩基対取得ステップを実行してから、図17に例示されるフローチャートを実行する。

[0068] (実施形態6:主に請求項6、7に対応する)

本発明の実施形態6として、塩基の不適合の発生の分布を指定して検索を行なう特異的塩基配列探索方法について説明する。

[0069] 本実施形態に係る特異的塩基配列探索方法は、実施形態2から5のいずれかに係る特異的塩基配列探索方法が、さらに、不適合分布情報取得ステップを含む方法である。

[0070] 「不適合分布情報取得ステップ」とは、塩基配列集合の塩基配列と特異的塩基配列候補が表わす塩基配列との適合の度合いとして、分布情報を取得する。「分布情報」とは、塩基の不適合の発生の分布を示す情報である。分布情報の例としては、塩基の不適合が連続して2以上存在しない、特異的塩基配列候補の5'末端側には不適合が少ない、特異的塩基配列候補との塩基の不適合が連続して所定の回数以上発生しない、などがある。このように分布情報を取得する目的としては、例えば、同じ数の塩基の不適合があっても、塩基の不適合が連続などしていると核酸がハイブリダイズしにくくなるので、適合許容数を満たしていても、塩基の不適合が連続などしてい

る塩基配列を排除することがある。また、塩基の不適合があっても、不適合とみなされない塩基が連続している場合には、不適合な部分があるにもかかわらず、ハイブリダイズする可能性が出てくるので、そのような場合を排除することを目的として、不適合とみなされない塩基が所定の値以上連続しないことを指定する。

[0071] 分布情報は、例えば、塩基の不適合の分布が所定の分布となっているかどうかを判定するプログラムであってもよい。あるいは、あらかじめ塩基の不適合の分布の類型をいくつか決めておき、それらを選択するための情報であってもよい。例えば、塩基の不適合の分布に番号をつけておき、その番号を示す情報であってもよい。

[0072] 本実施形態においては、塩基配列検索ステップの処理は、例えば、次のように行なう。すなわち、不適合分布情報取得ステップで取得された分布情報をさらに考慮に入れて、検索が行なわれる。例えば、まず、実施形態2から5のいずれかにおける検索を行ない、検索の結果から、不適合分布情報を満たすもの—例えば塩基の不適合が連続して2以上存在しないもの、特異的塩基配列候補の5′末端側に不適合が少ないもの、不適合とみなされない塩基が所定の数以上連続して発生しないもの—を選択する。

[0073] (実施形態7:主に請求項8に対応する)

本発明の実施形態7に係る特異的塩基配列探索方法は、実施形態2から6のいずれか—の特異的塩基配列探索方法において、特異的塩基配列候補を、マイクロアレイのオリゴDNAの塩基配列の候補とした方法である。

[0074] これにより、従来技術のように検索結果を精査する必要が無くなるので、マイクロアレイのオリゴDNAの設計を効率よく行なうことができる。

[0075] (実施形態8:主に請求項9に対応する)

本発明の実施形態8に係る特異的塩基配列探索方法は、実施形態2から6のいずれか—の特異的塩基配列探索方法において、特異的塩基配列候補を、siRNAの塩基配列の候補とした特異的塩基配列探索方法である。

[0076] これにより、従来技術のように検索結果を精査する必要が無くなるので、siRNAの配列の決定を効率よく行なうことができる。

[0077] (実施形態9:主に請求項10、11、21に対応する)



図19は、本発明の実施形態9に係る特異的塩基配列探索装置の機能ブロック図を例示する。本実施形態に係る特異的塩基配列探索装置は、例えば実施形態2に係る特異的塩基配列探索方法を使用するための装置である。

[0078] 特異的塩基配列探索装置1900は、塩基配列集合蓄積部1901と、特異的塩基配列候補取得部1902と、塩基配列検索部1903と、を有する。なお、本明細書において、機能ブロック図により表わされる構成は、ハードウェアとしては、任意の計算機のCPU、メモリ、その他のLSIなどにより実現される。また、ソフトウェアとしては、メモリにロードされたプログラムなどにより実現される。また、ハードウェアとソフトウェアとの連携により実現することもできる。特にソフトウェアが用いられて実現される場合には、これらの部は、計算機にプログラムをインストールして実行することにより実現可能である。例えば、プログラムは、各種の記録媒体に記録され、必要に応じて特異的塩基配列探索装置1900を実現するための計算機に機械的に読み取られる。ここで、「記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、EPROM、EEPROM、CD-ROM、MO、DVD、フラッシュディスク等の任意の「可搬用の物理媒体」や、各種計算機システムに内蔵されるROM、RAM、HD等の任意の「固定用の物理媒体」、あるいはLAN、WAN、インターネットに代表されるネットワークを介してプログラムを送信する場合の通信回線や搬送波のように短期にプログラムを保持する「通信媒体」を含むものとする。なお、ここにいう計算機とは、メインフレーム計算機に限定されることはなく、ワークステーションやパーソナルコンピュータなどの情報処理装置であってもよい。また、そのような情報処理装置には、プリンタやスキャナなどの周辺装置がされに接続されていてもよい。

[0079] また、「プログラム」とは、任意の言語や記述方法にて記述されたデータ処理方法であり、ソースコードやバイナリコード等の形式を問わない。なお、「プログラム」は必ずしも単一的に構成されるものに限られず、複数のモジュールやライブラリとして分散構成されるものや、オペレーティングシステムに代表される別個のプログラムと協同してその機能を達成するものをも含む。なお、特異的塩基配列探索装置1900において記録媒体を読み取るための具体的な構成、読み取り手段、あるいは、読み取り後のインストール手順等は、周知の構成や手順を用いることができる。



- [0080] ここでは図示を省略しているが、特異的塩基配列探索装置1900は、遺伝子の塩基配列情報等に関する外部データベースやホモロジ検索等の外部プログラム等を提供する外部システムに、インターネット等の通信網を介して通信可能に接続された構成であってもよい。かかる構成により、外部プログラムを実行するウェブサイトが提供される。外部システムは、WEBサーバやASPサーバ等として構成されてもよい。例えば、塩基配列集合蓄積部1901及び／又は特異的塩基配列候補取得部1902が外部システムに通信可能に接続されてもよい。通信網の構成は特には限定されないが、例えば、ルータ等の通信装置や専用線等の有線又は無線の通信回線により構成される。
- [0081] 「塩基配列集合蓄積部」1901は、塩基配列集合を保持する。「塩基配列集合」とは、エクソンの塩基配列の和集合と、複数のエクソンから構成される発現遺伝子におけるエクソンの境界にまたがって存在する境界塩基配列の集合と、の和集合を含む集合である。例えば、実施形態1で説明した方法により生成された集合である。あるいは、実施形態2の方法などの塩基配列検索ステップにて検索がされる集合である。塩基配列集合蓄積部1901は、塩基配列集合を、例えば、RAM、ROMなどのメモリ装置、ハードディスクなどの固定ディスク装置、フレキシブルディスクや光ディスクを用いる蓄積装置を用いて所定のフォーマット、形式としてのデータとして入出力可能に蓄積する。したがって、特異的塩基配列探索装置1900が計算機を用いて実現される場合には、これら蓄積のための装置との入出力を行なうドライバ及び、そのドライバを用いてデータの入出力を行なうプログラムのモジュールなどが、塩基配列集合蓄積部1901に相当する。
- [0082] 「特異的塩基配列候補取得部」1902は、発現遺伝子の塩基配列に特異的に現れる塩基配列の候補である特異的塩基配列候補を取得する。例えば、インターネットなどの通信網によって通信を行なうことができる計算機で動作するウェブブラウザに表示されたウェブページのテキストエリアに入力され、HTTP(HyperText Transfer Protocol)を用いてそのブラウザからテキスト情報などとして送信された特異的塩基配列候補を受信することにより、特異的塩基配列候補の取得がされる。したがって、特異的塩基配列探索装置1900が計算機を用いて実現される場合には、通信インタ

ーフースや、マウス、キーボード、ディスプレイとのデータの入出力を行なうための入出力インターフェースにおける入出力を行なうドライバ、及び、そのドライバを用いてデータの入出力を行なうプログラムのモジュールなどが、特異的塩基配列候補取得部1902に相当する。

[0083] 「塩基配列検索部」1903は、塩基配列集合蓄積部1901に蓄積された塩基配列集合に含まれる塩基配列から、特異的塩基配列候補取得部1902で取得された特異的塩基配列候補と適合する塩基配列である適合塩基配列を検索する。この検索には、例えば、実施形態2から4のいずれかで説明したアルゴリズム(例えば、BLAST)を実行するプログラムを用いる。検索の結果は、特異的塩基配列候補を送信したブラウザに返信するようになっていてもよい。例えば、検索の結果の件数を返信したり、特異的塩基配列候補に適合する塩基配列を発現遺伝子に関する情報を取得して返信を行なったりしてもよい。また、検索の結果の件数に応じて、特異的塩基配列候補取得部1902が取得した特異的塩基配列候補が特異的塩基配列であるかどうかを判断した結果を返信するようになっていてもよい。また、ブラウザの内部でJAVA(登録商標)などで記述されたプログラムが動作しており、そのプログラムにより、特異的塩基配列候補が特異的塩基配列であるかどうかを判断するようになっていてもよい。なお、特異的塩基配列探索装置1900が計算機を用いて実現される場合には、計算機のCPUによる制御の下で、特異的塩基配列候補取得部1902に相当するモジュールなどとデータの受け渡しを行ない、また、塩基配列集合蓄積部1901に相当するモジュールなどとデータの受け渡しを行ない、ハードディスクなどに蓄積された塩基配列集合の検索を行なうモジュールなどが、塩基配列検索部1903に相当する。

[0084] また、特異的塩基配列探索装置1900には、塩基配列検索部1903による検索の結果を蓄積する部が備わっていてもよい。特異的塩基配列候補取得部1902で取得された特異的塩基配列候補と、塩基配列検索部1903による検索の結果と、を関連付けて蓄積する部が備わっていてもよい。このような部を備えることにより、同じ特異的塩基配列候補が特異的塩基配列候補取得部1902で複数回取得される場合には、二回目以降は、この部に蓄積された情報を検索することにより、応答スピードを速くすることができる。

[0085] (実施形態10:主に請求項12に対応する)

本発明の実施形態10は、実施形態9の特異的塩基配列探索装置において、境界塩基配列集合を、複数のエクソンから構成される発現遺伝子におけるエクソンの境界にまたがって存在する塩基配列を示す情報であって、前記特異的塩基配列候補の塩基配列の長さと同じ長さの塩基配列を示す情報、からなる集合に対して、発現遺伝子が同じで、塩基配列の位置が重複する塩基配列を示す情報を統合することにより得られる集合に基づいて得られるものとした特異的塩基配列探索装置である。本実施形態に係る特異的塩基配列探索装置は、例えば実施形態3に係る特異的塩基配列探索方法を使用するための装置である。

[0086] すなわち、本実施形態に係る特異的塩基配列探索装置は、塩基配列集合蓄積部1901に蓄積される塩基配列集合を、境界塩基配列に第七節などで説明した処理による統合の処理を行なった集合とした特異的塩基配列探索装置である。

[0087] 統合を行なうことにより、塩基配列集合の要素数を減少させることができるので、塩基配列集合蓄積部1901が使用するディスクスペースを節約することができる。また、要素数の減少による検索速度の向上も実現される。

[0088] (実施形態11:主に請求項13に対応する)

図20は、本発明の実施形態11に係る特異的塩基配列探索装置の機能ブロック図を例示する。特異的塩基配列探索装置2000は、塩基配列集合蓄積部1901と、特異的塩基配列候補取得部1902と、塩基配列検索部1903と、適合許容数取得部2001と、を有する。したがって、本実施形態に係る特異的塩基配列探索装置は、実施形態9または10に係る特異的塩基配列探索装置が適合許容数取得部を有した構成となっている。なお、本明細書においては、同じ定義が適用できる部には、同じ符号を割り当てることとする。ただし、実際の製造においては、同じ符号が割り当てられているからといって、作りなどが同じになるとは限らない。なお、本実施形態に係る特異的塩基配列探索装置は、例えば実施形態4に係る特異的塩基配列探索方法を使用するための装置である。

[0089] 「適合許容数取得部」2001は、塩基配列集合の塩基配列と特異的塩基配列候補が表わす塩基配列との適合の度合いとして、いくつかの塩基の不適合まで許容するか

を示す数値である適合許容数を取得する。例えば、特異的塩基配列候補がブラウザから送信されるときに、そのブラウザから適合許容数も送信されてもよい。適合許容数取得部2001は、そのように送信される適合許容数を取得する。また、適合許容数を直接入力する構成であってもよい。

[0090] 本実施形態においては、塩基配列検索部1903は、適合許容数取得部2001にて取得された適合許容数に基づいて検索を行なう。この検索の方法については、実施形態4で述べたとおりである。

[0091] (実施形態12:主に請求項14に対応する)

図21は、本発明の実施形態12に係る特異的塩基配列探索装置の機能ブロック図を例示する。特異的塩基配列探索装置2100は、塩基配列集合蓄積部1901と、特異的塩基配列候補取得部1902と、塩基配列検索部1903と、適合許容数取得部2001と、不適合塩基対取得部2101と、を有する。したがって、本実施形態に係る特異的塩基配列探索装置は、実施形態11に係る特異的塩基配列探索装置が、さらに、不適合塩基対取得部2101を有する構成となっている。本実施形態に係る特異的塩基配列探索装置は、例えば実施形態5に係る特異的塩基配列探索方法を使用するための装置である。

[0092] 「不適合塩基対取得部」2101は、塩基配列検索部による検索において、不適合と判断する塩基の対を取得する。例えば、不適合と判断すべき塩基の対を示すテキスト情報を取得する。あるいは、適合と判断すべき塩基の対(例えば、GとU)を取得することにより、間接的に不適合と判断すべき塩基の対を取得するようになっていてもよい。したがって、通信インターフェースや、マウス、キーボード、ディスプレイとのデータの入出力を行なうための入出力インターフェースにおける入出力を行なうドライバ、及び、そのドライバを用いてデータの入出力を行なうプログラムのモジュールなどが不適合塩基対取得部2101に相当する。

[0093] 本実施形態に係る特異的塩基配列探索装置の処理の流れは、実施形態11に係る特異的塩基配列探索装置と同じである。ただし、適合塩基配列を検索する前に、不適合塩基対取得部2101により塩基配列検索部による検索において、不適合と判断する塩基の対を取得することが行なわれる。



[0094] (実施形態13:主に請求項15、16に対応する)

図22は、本発明の実施形態13に係る特異的塩基配列探索装置の機能ブロック図を例示する。特異的塩基配列探索装置2200は、塩基配列集合蓄積部1901と、特異的塩基配列候補取得部1902と、塩基配列検索部1903と、適合許容数取得部2001と、不適合分布情報取得部2201と、を有する。また、特異的塩基配列探索装置2200は、さらに、不適合塩基対取得部を有していてもよい。したがって、本実施形態に係る特異的塩基配列探索装置は、実施形態9から12のいずれかに係る特異的塩基配列探索装置が、不適合分布情報取得部2201を有した構成となっている。本実施形態に係る特異的塩基配列探索装置は、例えば実施形態6に係る特異的塩基配列探索方法を使用するための装置である。

[0095] 「不適合分布情報取得部」2201は、塩基配列集合の塩基配列と特異的塩基配列候補が表わす塩基配列との適合の度合いとして、塩基の不適合の発生の分布を示す情報である分布情報を取得する。分布情報の例としては、実施形態6で述べたとおりである。したがって、通信インターフェースや、マウス、キーボード、ディスプレイとのデータの入出力を行なうための入出力インターフェースにおける入出力を行なうドライバ、及び、そのドライバを用いてデータの入出力を行なうプログラムのモジュールなどが不適合分布情報取得部2201に相当する。

[0096] 本実施形態においては、塩基配列検索部1903は、不適合分布情報取得部2201で取得された分布情報に基づいて検索を行なう。例えば、実施形態11または実施形態12におけるように検索を行ない、その検索の結果である中間結果から、分布情報に基づいて検索を行なう。すなわち、中間結果から、分布情報に合致するものを選択して、最終的な検索の結果とする。

[0097] (実施形態14:主に請求項17、18に対応する)

本発明の実施形態14は、塩基配列集合保持装置である。すなわち、エクソンの塩基配列の和集合と、複数のエクソンから構成される発現遺伝子におけるエクソンの境界にまたがって存在する塩基配列である境界塩基配列の集合と、の和集合を含む集合である塩基配列集合を、検索可能に保持する装置である。

[0098] したがって、本実施形態に係る塩基配列集合保持装置の形態としては、例えば実



施形態8に係る特異的塩基配列探索装置1900の塩基配列集合蓄積部1901を実現するハードディスクが外付けハードディスク装置になっているものを挙げることができる。また、特異的塩基配列探索装置1900の塩基配列集合蓄積部1901を実現するハードディスクを有するサーバ装置であってもよい。

[0099] 本実施形態に係る塩基配列集合保持装置により、様々な検索アルゴリズムに基づく検索を実現することが可能となる。

[0100] (実施形態15:主に請求項19、22に対応する)

本発明の実施形態15は、実施形態14の塩基配列集合保持装置において、保持される境界塩基配列の集合を、複数のエクソンから構成される発現遺伝子におけるエクソンの境界にまたがって存在する塩基配列を示す情報であって、検索の入力となる塩基配列の長さと同じ長さの塩基配列を示す情報、からなる集合に対して、発現遺伝子が同じで、塩基配列の位置が重複する塩基配列を示す情報を統合することにより得られる集合に基づいて得られるものとした、塩基配列集合保持装置である。すなわち、実施形態10に係る特異的塩基配列探索装置の塩基配列集合蓄積部を、別の装置とした形態である。例えば、実施形態10に係る特異的塩基配列探索装置の塩基配列集合蓄積部に蓄積されるデータを、NAS(Network Attached Storage)やSAN(Storage Area Network)に蓄積させることにより得られる形態である。

[0101] 本実施形態においては、境界塩基配列に対して統合の操作が行なわれるので、必要なディスクスペースを減少させることができる。

### 産業上の利用可能性

[0102] 本発明においては、エクソンの塩基配列と、エクソンの境界に現れる塩基配列と、から塩基配列集合を生成して、検索を行なうので、発現遺伝子に特異的に現れる塩基配列かどうかを検索結果数に基づいて決定できるので、特異的塩基配列を決定する上で有用である。

### 図面の簡単な説明

[0103] [図1]RNA干渉の過程の概略を示す図

[図2]マイクロアレイを用いた遺伝子解析や遺伝子診断などの過程の一例図

[図3]DNA配列とmRNAに転写される発現遺伝子配列との関係の一例図

[図4]エクソンの和集合と発現遺伝子のエクソンの境界をまたぐ塩基配列の一例図

[図5]N-1通りの境界塩基配列の一例図

[図6]塩基配列の統合を説明するための図

[図7]塩基配列の統合を説明するための図

[図8]塩基配列の和集合を計算するために用いるテーブルの一例図

[図9]塩基配列の和集合を計算するためのフローチャート

[図10]境界塩基配列の統合の求め方の一例図

[図11]長さがN-1mer未満のエクソンが存在する場合の一例図

[図12]統合の操作を行なうために使用するテーブルの一例図

[図13]統合の処理のフローチャート

[図14]本発明の実施形態1に係る塩基配列集合生成方法の処理のフローチャート

[図15]和集合生成ステップで得られた塩基配列を格納したテーブルの一例図

[図16]本発明の実施形態2に係る特異的塩基配列探索方法のフローチャート

[図17]本発明の実施形態4に係る特異的塩基配列探索方法のフローチャート

[図18]候補塩基配列長が19で適合許容数が3である場合にBLASTでは検索できないと考えられる塩基配列のミスマッチを示す図

[図19]本発明の実施形態9に係る特異的塩基配列探索装置の機能ブロック図

[図20]本発明の実施形態11に係る特異的塩基配列探索装置の機能ブロック図

[図21]本発明の実施形態12に係る特異的塩基配列探索装置の機能ブロック図

[図22]本発明の実施形態13に係る特異的塩基配列探索装置の機能ブロック図

## 符号の説明

[0104]	301	エクソン
	302	エクソン
	303	エクソン
	304	エクソン
	305	エクソン
	306	エクソン
	401	エクソンの和集合の一要素

402 エクソンの和集合の一要素

403 エクソンの和集合の一要素

404 境界塩基配列の一部

405 境界塩基配列の一部

406 境界塩基配列の一部

407 境界塩基配列の一部

## 請求の範囲

- [1] 発現遺伝子の塩基配列に特異的に現れる塩基配列の候補である特異的塩基配列候補を取得する特異的塩基配列候補取得ステップと、  
エクソンの塩基配列の和集合と、  
複数のエクソンから構成される発現遺伝子におけるエクソンの境界にまたがって存在する塩基配列である境界塩基配列の集合と、  
の和集合を含む集合である塩基配列集合の中から、前記特異的塩基配列候補取得ステップにて取得された特異的塩基配列候補が表わす塩基配列と適合する塩基配列である適合塩基配列を検索する塩基配列検索ステップと、  
前記塩基配列検索ステップでの検索結果に適合塩基配列が複数あるかどうかに基づいて、前記特異的塩基配列候補取得ステップにて取得された特異的塩基配列候補が特異的塩基配列であるか判断する判断ステップと、  
を含む特異的塩基配列探索方法。
- [2] 前記エクソンの塩基配列の和集合の要素には、エクソンの配列位置を示す情報またはエクソンが構成する遺伝子を識別する情報を含む属性情報が関連付けられている請求項1に記載の特異的塩基配列探索方法。
- [3] 前記境界塩基配列の集合は、  
複数のエクソンから構成される発現遺伝子におけるエクソンの境界にまたがって存在する塩基配列を示す情報であって、前記特異的塩基配列候補の塩基配列の長さと同じ長さの塩基配列を示す情報、からなる集合に対して、発現遺伝子が同じで、塩基配列の位置が重複する塩基配列を示す情報を統合することにより得られる集合に基づいて得られるものである請求項1または2に記載の特異的塩基配列探索方法。
- [4] 前記塩基配列集合の塩基配列と前記特異的塩基配列候補が表わす塩基配列との適合の度合いとして、いくつかの塩基の不適合まで許容するかを示す数値である適合許容数を取得する適合許容数取得ステップを含み、  
前記塩基配列検索ステップでは、前記適合許容数取得ステップにて取得された適合許容数に基づいて検索を行なう請求項1から3のいずれか一に記載の特異的塩基配列探索方法。



- [5] 前記塩基配列検索ステップにおいて不適合と判断する塩基の対を取得する不適合塩基対取得ステップを含む請求項4に記載の特異的塩基配列探索方法。
- [6] 前記塩基配列集合の塩基配列と前記特異的塩基配列候補が表わす塩基配列との適合の度合いとして、塩基の不適合の発生の分布を示す情報である分布情報を取得する不適合分布情報取得ステップを含み、  
前記塩基配列検索ステップでは、前記不適合分布情報取得ステップで取得された分布情報に基づいて検索を行なう請求項1から5のいずれかーに記載の特異的塩基配列探索方法。
- [7] 前記分布情報は、不適合とみなされない塩基が連続する長さを表わす請求項6に記載の特異的塩基配列探索方法。
- [8] 前記特異的塩基配列候補は、マイクロアレイのオリゴDNAの塩基配列の候補である請求項1から7のいずれかーに記載の特異的塩基配列探索方法。
- [9] 前記特異的塩基配列候補は、siRNAの塩基配列の候補を示す請求項1から7のいずれかーに記載の特異的塩基配列探索方法。
- [10] エクソンの塩基配列の和集合と、  
複数のエクソンから構成される発現遺伝子におけるエクソンの境界にまたがって存在する塩基配列である境界塩基配列の集合と、  
の和集合を含む集合である塩基配列集合を保持する塩基配列集合蓄積部と、  
発現遺伝子の塩基配列に特異的に現れる塩基配列の候補である特異的塩基配列候補を取得する特異的塩基配列候補取得部と、  
前記塩基配列集合蓄積部に蓄積された塩基配列集合に含まれる塩基配列から、前記特異的塩基配列候補取得部で取得された特異的塩基配列候補と適合する塩基配列である適合塩基配列を検索する塩基配列検索部と、  
を有する特異的塩基配列探索装置。
- [11] 前記エクソンの塩基配列の和集合の要素には、エクソンの配列位置を示す情報またはエクソンが構成する遺伝子を識別する情報を含む属性情報が関連付けられている請求項10に記載の特異的塩基配列探索装置。
- [12] 前記境界塩基配列の集合は、

複数のエクソンから構成される発現遺伝子におけるエクソンの境界にまたがって存在する塩基配列を示す情報であって、前記特異的塩基配列候補の塩基配列の長さと同じ長さの塩基配列を示す情報、からなる集合に対して、発現遺伝子が同じで、塩基配列の位置が重複する塩基配列を示す情報を統合することにより得られる集合に基づいて得られるものである請求項10または11に記載の特異的塩基配列探索装置。

- [13] 前記塩基配列集合の塩基配列と前記特異的塩基配列候補が表わす塩基配列との適合の度合いとして、いくつかの塩基の不適合まで許容するかを示す数値である適合許容数を取得する適合許容数取得部を有し、

前記塩基配列検索部は、前記適合許容数取得部にて取得された適合許容数に基づいて検索を行なう請求項10から12のいずれか一に記載の特異的塩基配列探索装置。

- [14] 前記塩基配列検索部による検索において不適合と判断する塩基の対を取得する不適合塩基対取得部を有する請求項13に記載の特異的塩基配列探索装置。

- [15] 前記塩基配列集合の塩基配列と前記特異的塩基配列候補が表わす塩基配列との適合の度合いとして、塩基の不適合の発生の分布を示す情報である分布情報を取得する不適合分布情報取得部を有し、

前記塩基配列検索部は、前記不適合分布情報取得部で取得された分布情報に基づいて検索を行なう請求項10から14のいずれか一に記載の特異的塩基配列探索装置。

- [16] 前記分布情報は、不適合とみなされない塩基が連続する長さを表わす請求項15に記載の特異的塩基配列探索装置。

- [17] エクソンの塩基配列の和集合と、

複数のエクソンから構成される発現遺伝子におけるエクソンの境界にまたがって存在する塩基配列である境界塩基配列の集合と、

の和集合を含む集合である塩基配列集合を、検索可能に保持する塩基配列集合保持装置。

- [18] 前記エクソンの塩基配列の和集合の要素には、エクソンの配列位置を示す情報ま

たはエクソンが構成する遺伝子を識別する情報を含む属性情報が関連付けられている請求項17に記載の塩基配列集合保持装置。

[19] 前記境界塩基配列の集合は、

複数のエクソンから構成される発現遺伝子におけるエクソンの境界にまたがって存在する塩基配列を示す情報であって、検索の入力となる塩基配列の長さと同じ長さの塩基配列を示す情報、からなる集合に対して、発現遺伝子が同じで、塩基配列の位置が重複する塩基配列を示す情報を統合することにより得られる集合に基づいて得られるものである請求項17または18に記載の塩基配列集合保持装置。

[20] 発現遺伝子の塩基配列に特異的に現れる塩基配列の候補の塩基配列の長さを取得する候補塩基配列長取得ステップと、

エクソンの塩基配列の和集合を取得するエクソン塩基配列集合取得ステップと、

複数のエクソンから構成される発現遺伝子におけるエクソンの境界にまたがって存在する塩基配列を示す情報であって、候補塩基配列長ステップで取得された長さと同じ長さの塩基配列を示す情報、からなる集合に対して、発現遺伝子が同じで、塩基配列の位置が重複する塩基配列を示す情報を統合することにより塩基配列の集合を生成する境界塩基配列集合生成ステップと、

前記エクソン塩基配列集合取得ステップで取得された塩基配列の集合と、前記境界塩基配列集合生成ステップで生成された塩基配列の集合との和集合を生成する和集合生成ステップと、

を含む塩基配列集合生成方法。

[21] 発現遺伝子の塩基配列に特異的に現れる塩基配列の候補である特異的塩基配列候補を取得する特異的塩基配列候補取得ステップと、

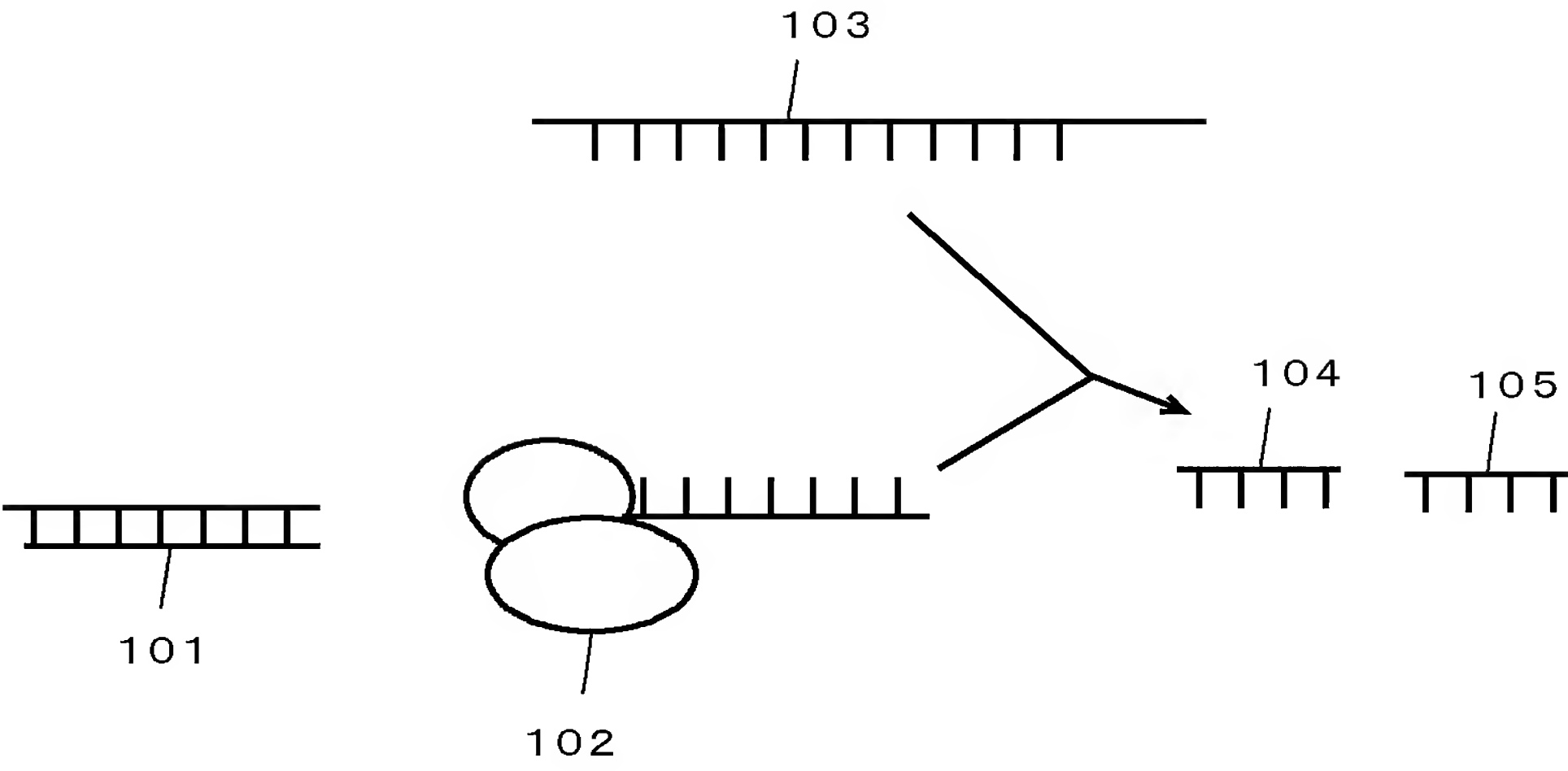
エクソンの塩基配列の和集合と、複数のエクソンから構成される発現遺伝子におけるエクソンの境界にまたがって存在する塩基配列である境界塩基配列の集合と、の和集合を含む集合である塩基配列集合に含まれる塩基配列から、前記特異的塩基配列候補取得ステップで取得された特異的塩基配列候補と適合する塩基配列である適合塩基配列を検索する塩基配列検索ステップと、

を計算機に実行させるための特異的塩基配列探索プログラム。

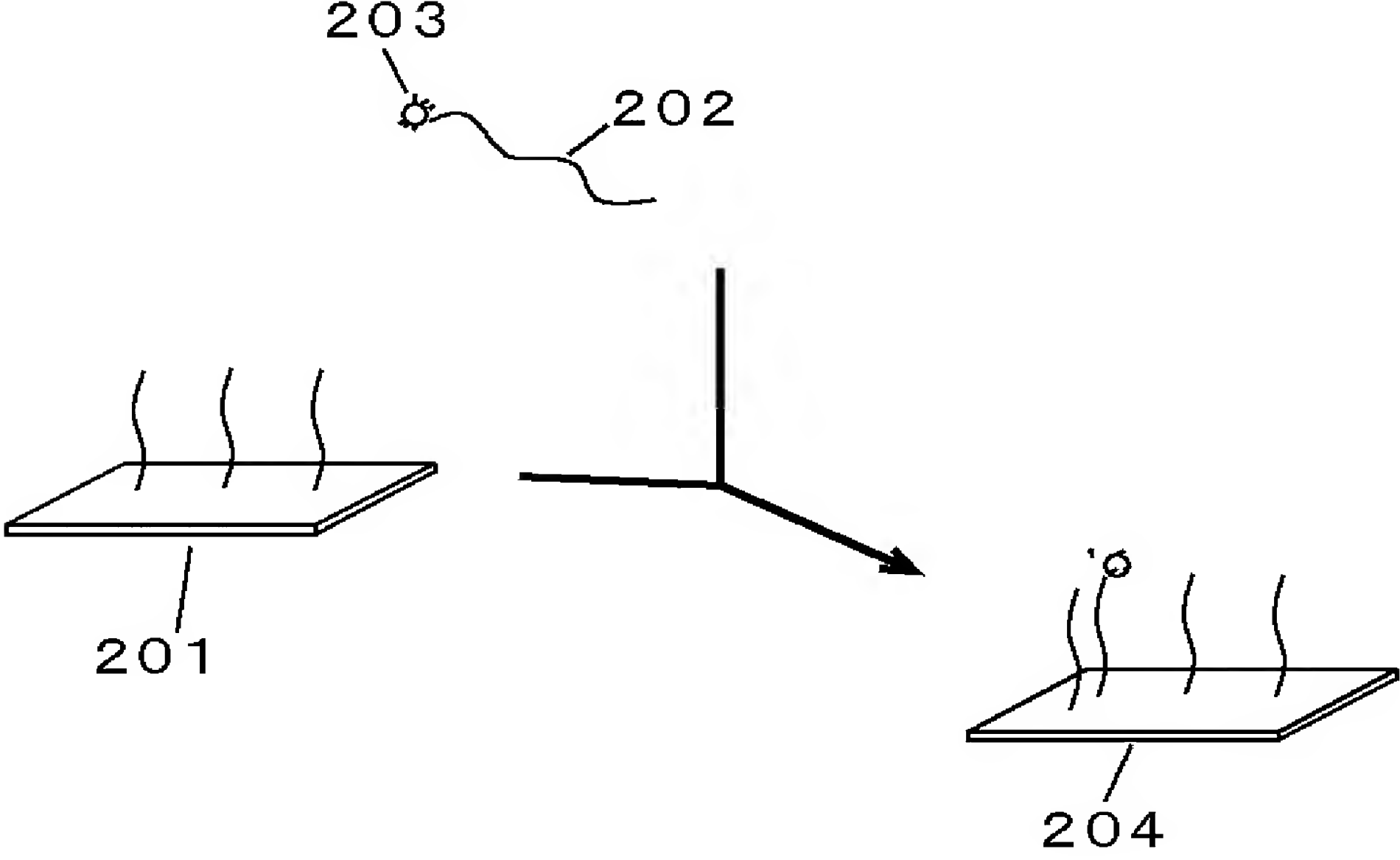
- [22] 発現遺伝子の塩基配列に特異的に現れる塩基配列の候補の塩基配列の長さを取得する候補塩基配列長取得ステップと、  
エクソンの塩基配列の和集合を取得するエクソン塩基配列集合取得ステップと、  
複数のエクソンから構成される発現遺伝子におけるエクソンの境界にまたがって存在する塩基配列を示す情報であって、候補塩基配列長ステップで取得された長さと同じ長さの塩基配列を示す情報、からなる集合に対して、発現遺伝子と同じで、塩基配列の位置が重複する塩基配列を示す情報を統合することにより塩基配列の集合を生成する境界塩基配列集合生成ステップと、  
前記エクソン塩基配列集合取得ステップで取得された塩基配列の集合と、前記境界塩基配列集合生成ステップで生成された塩基配列の集合との和集合を生成する和集合生成ステップと、  
を計算機に実行させるための塩基配列集合生成プログラム。
- [23] 発現遺伝子の塩基配列に特異的に現れる塩基配列の候補である特異的塩基配列候補を取得する特異的塩基配列候補取得ステップと、  
エクソンの塩基配列の和集合と、  
複数のエクソンから構成される発現遺伝子におけるエクソンの境界にまたがって存在する塩基配列である境界塩基配列の集合と、  
の和集合を含む集合である塩基配列集合の中から、前記特異的塩基配列候補取得ステップにて取得された特異的塩基配列候補が表わす塩基配列と適合する塩基配列である適合塩基配列を検索する塩基配列検索ステップと、  
前記塩基配列検索ステップでの検索結果に適合塩基配列が複数あるかどうかに基づいて、前記特異的塩基配列候補取得ステップにて取得された特異的塩基配列候補が特異的塩基配列であるか判断する判断ステップと、  
を計算機に実行させるための特異的塩基配列探索プログラム。



[図1]

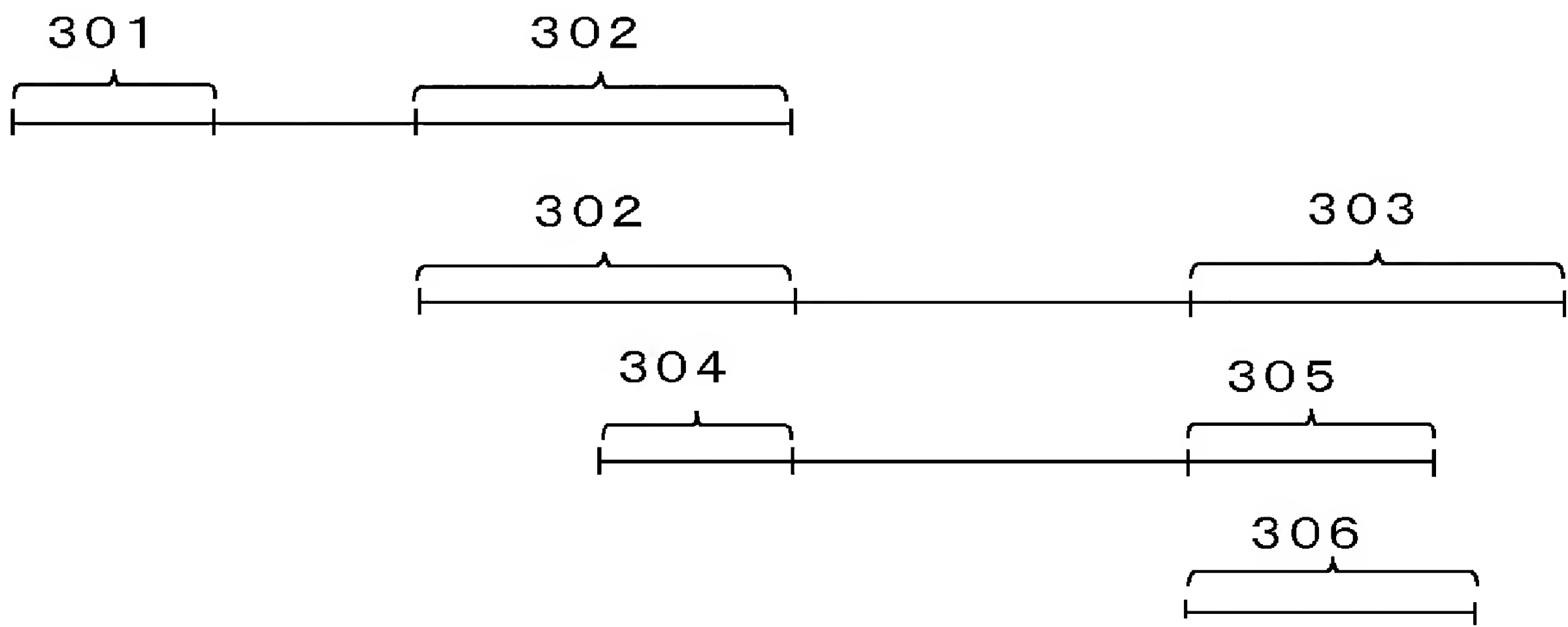


[図2]

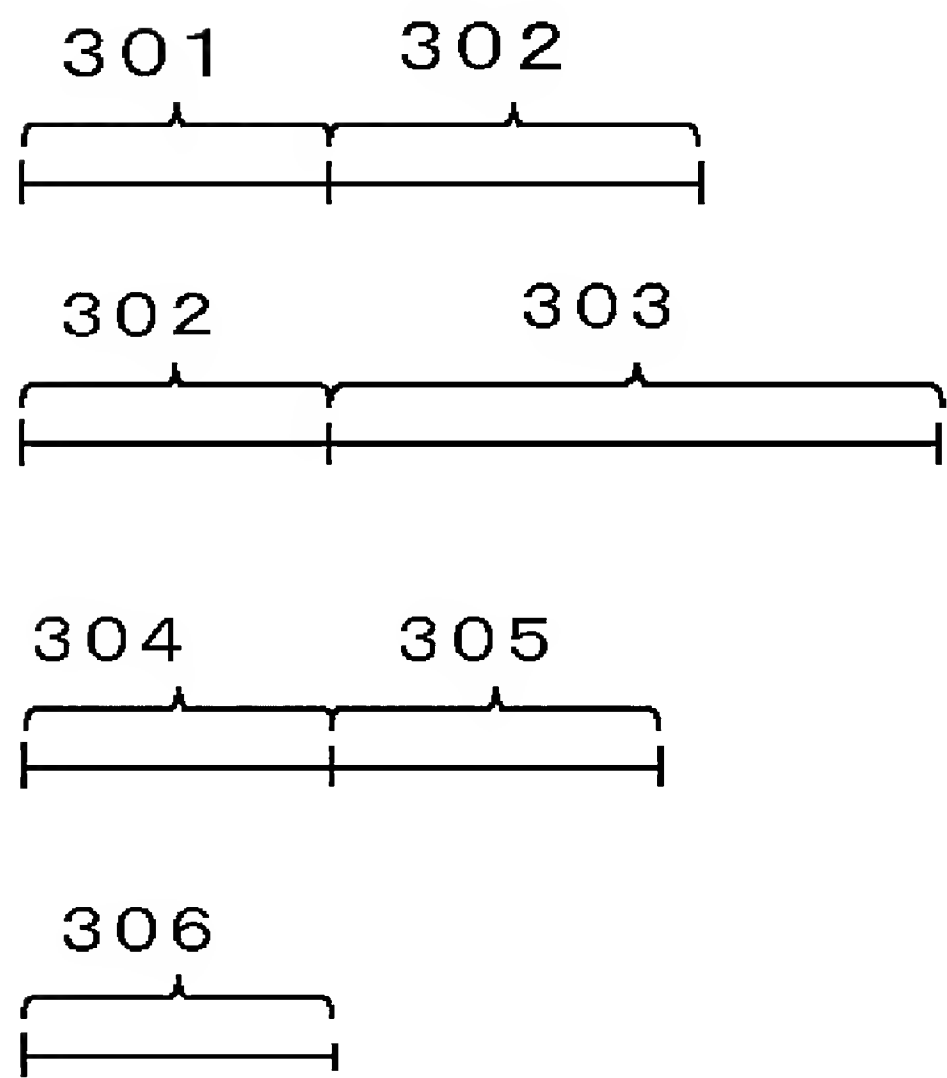


[図3]

(A)

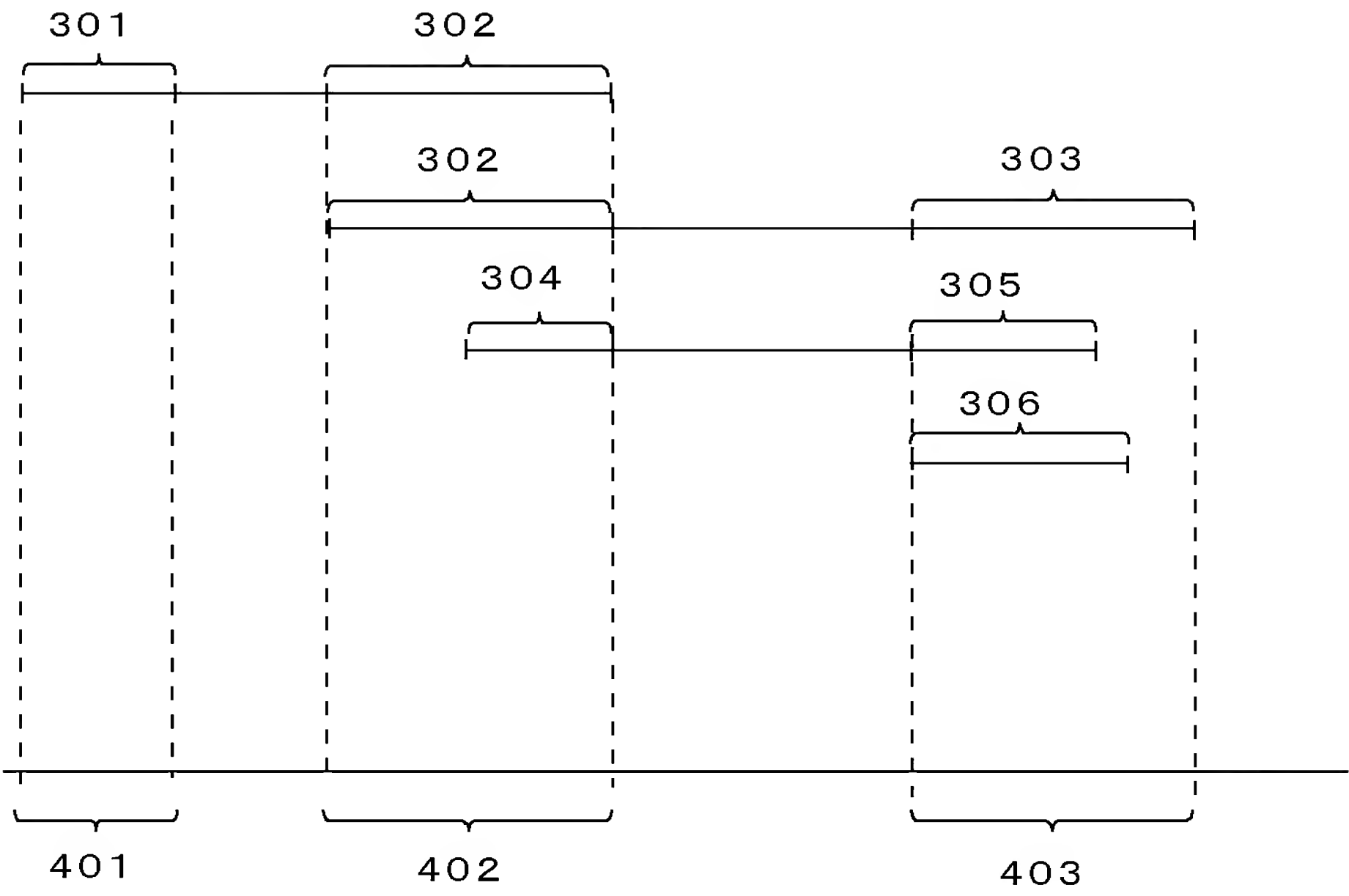


(B)

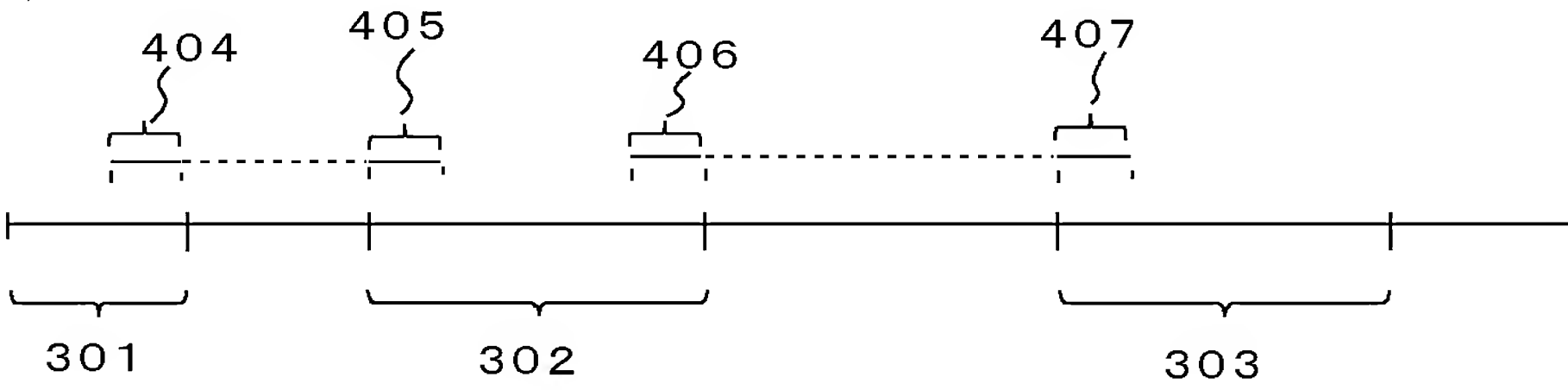


[図4]

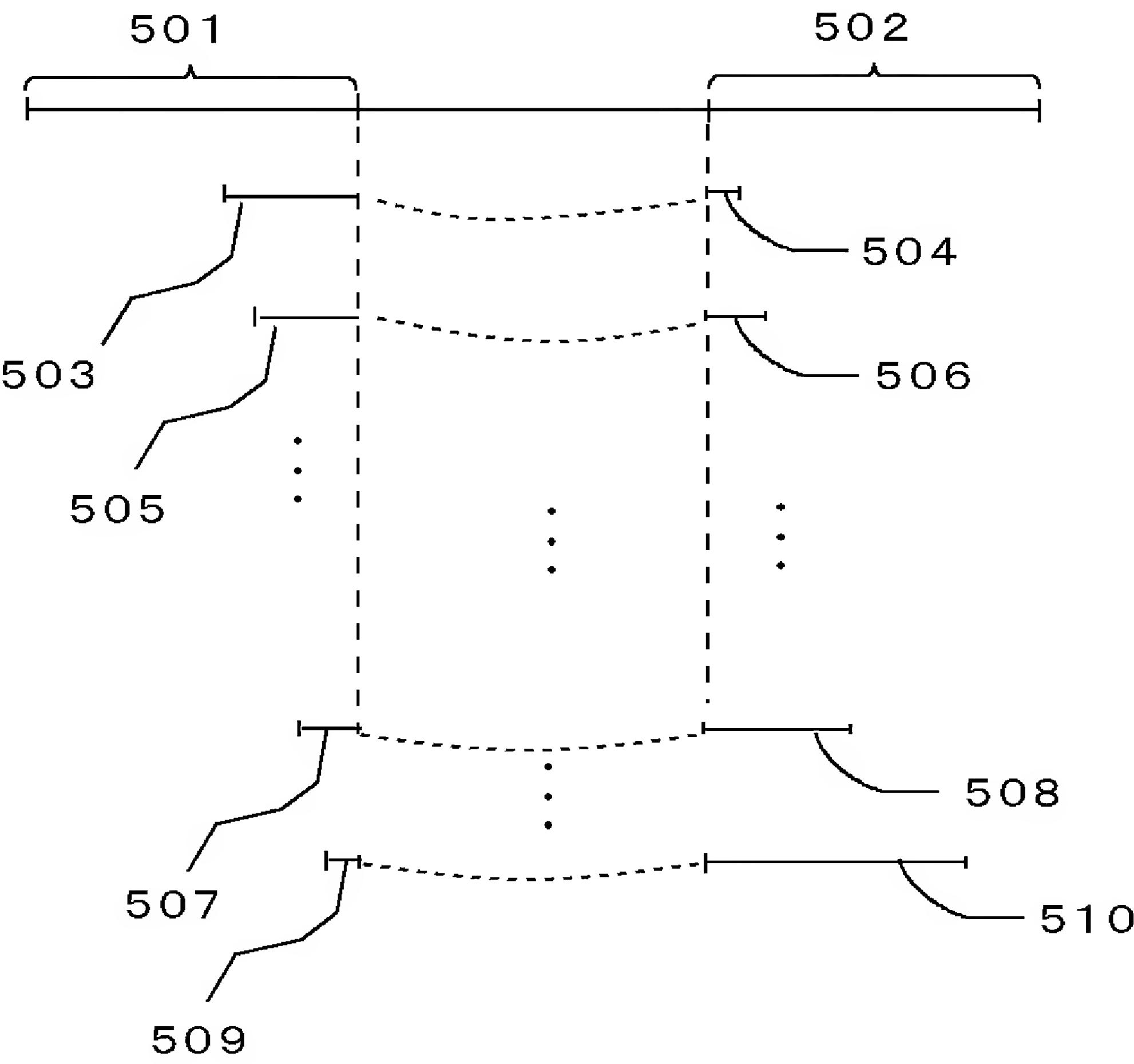
(A)



(B)

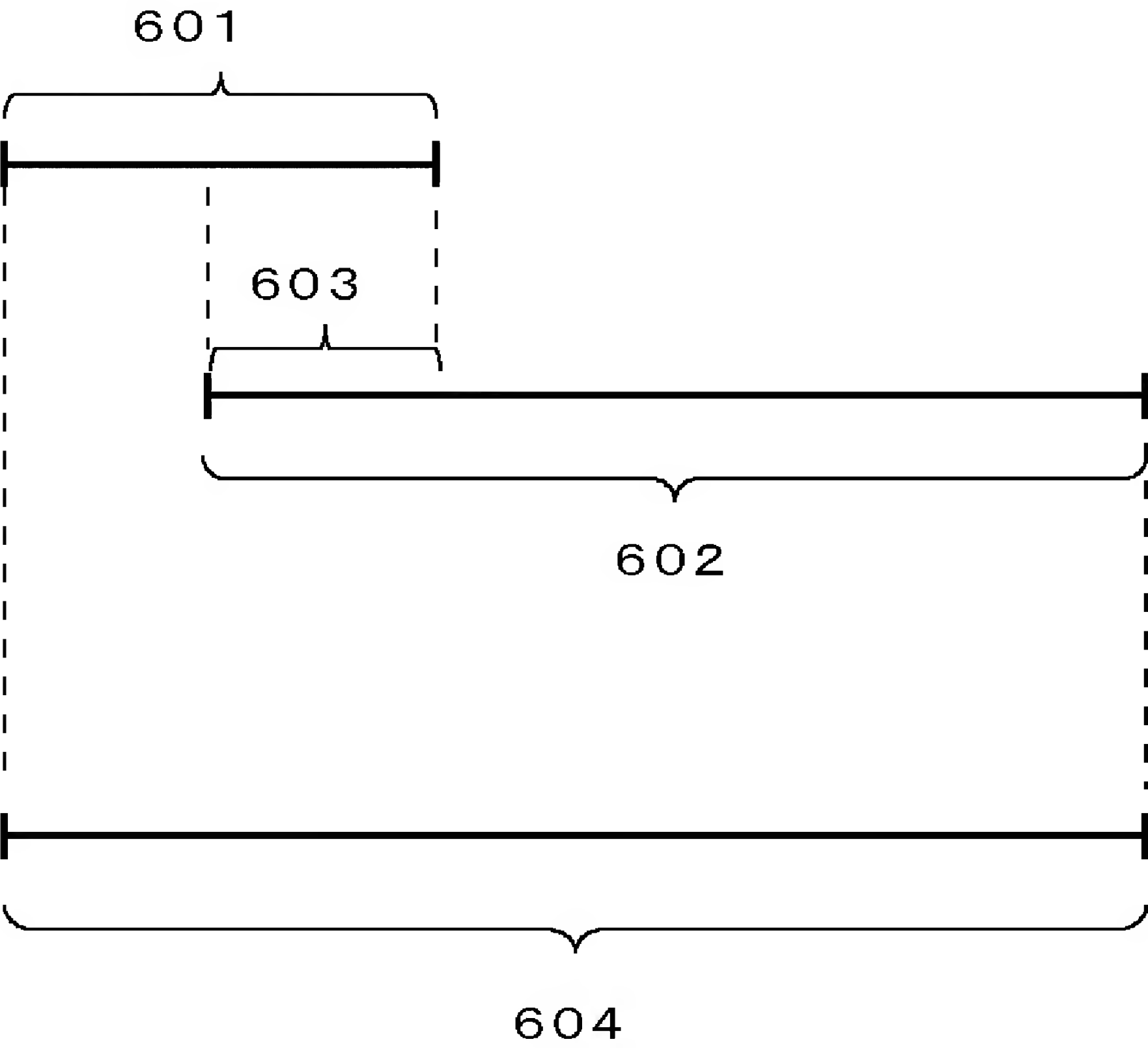


[図5]

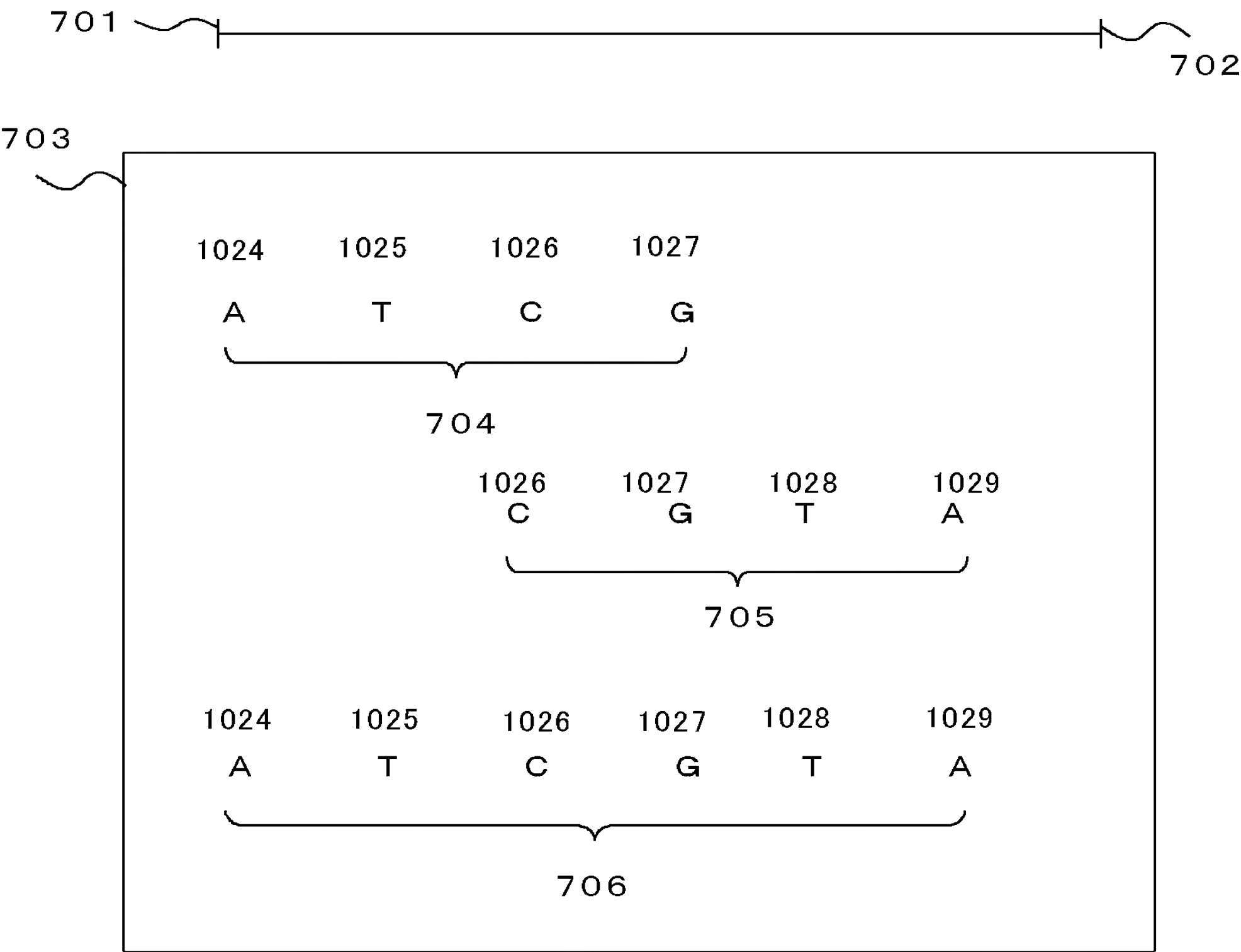




[図6]



[図7]



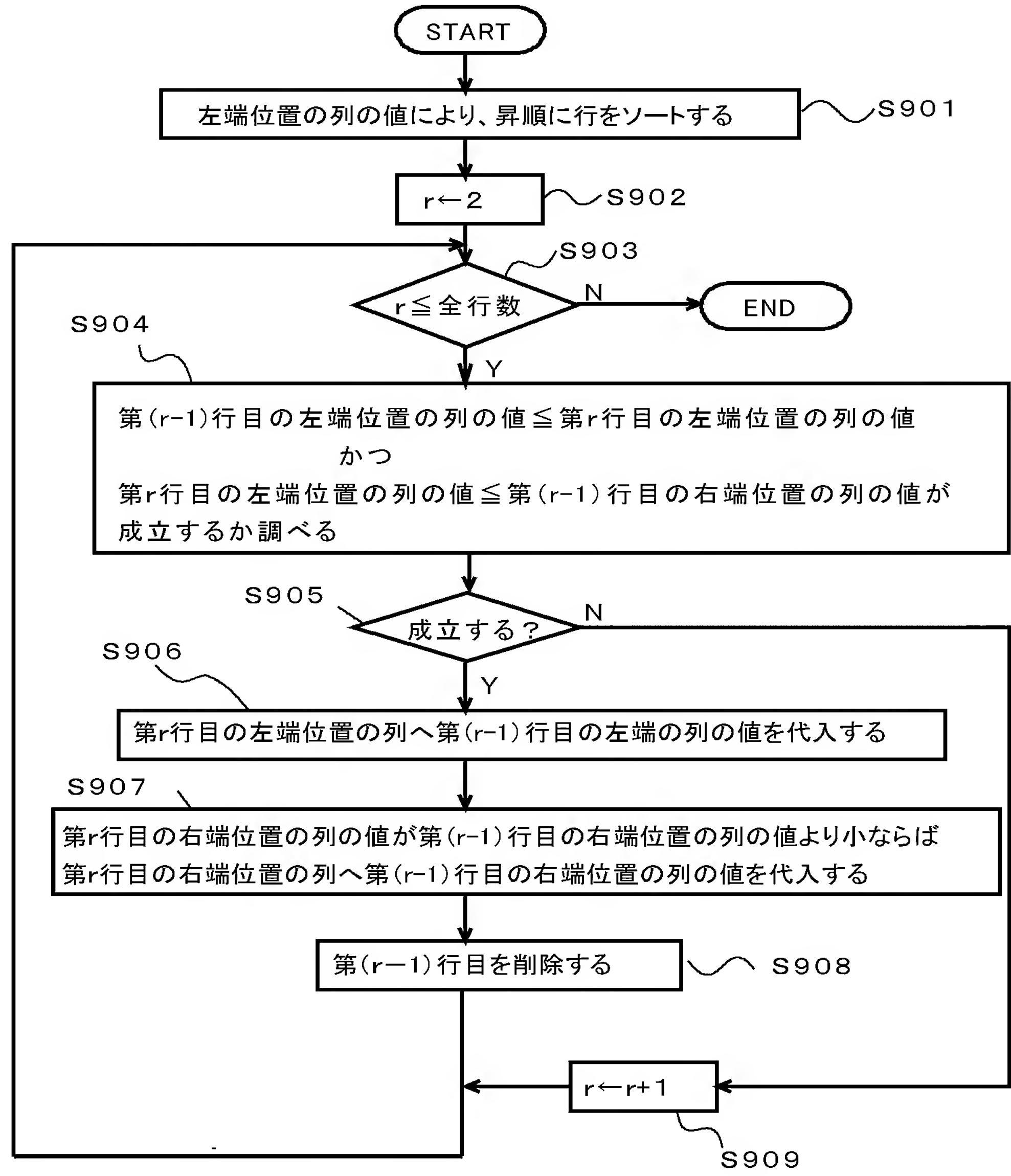
[図8]

801

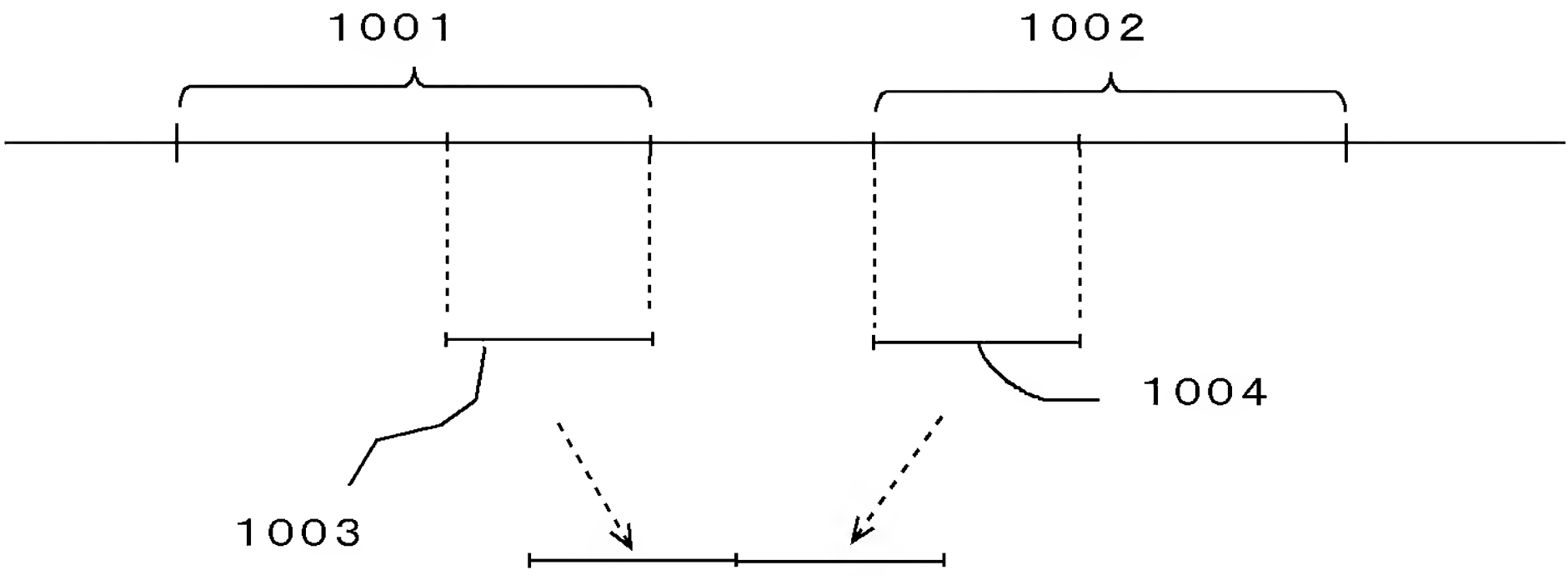
802

左端位置	右端位置
1024	1027
1026	1029
⋮	⋮

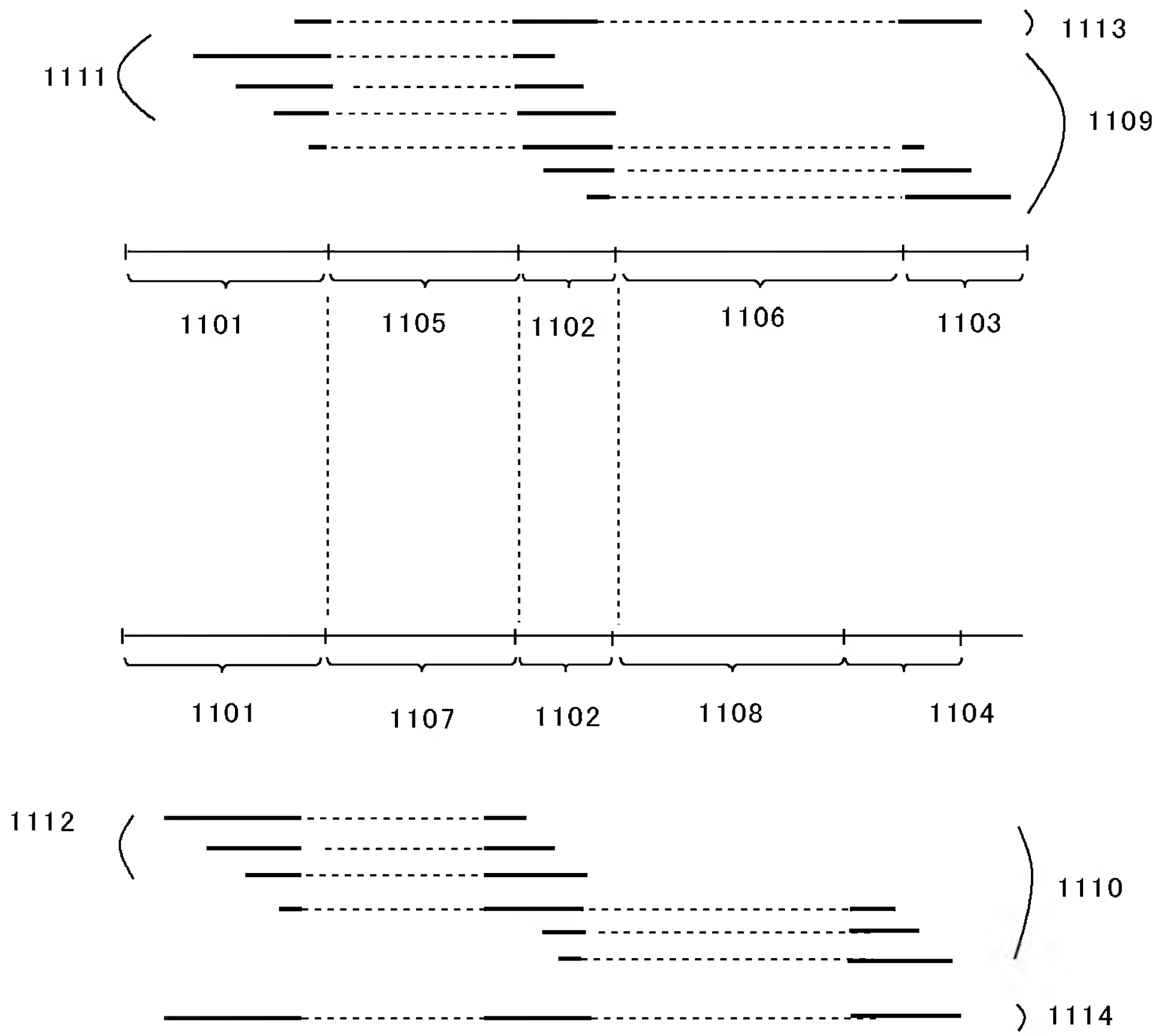
[図9]



[図10]



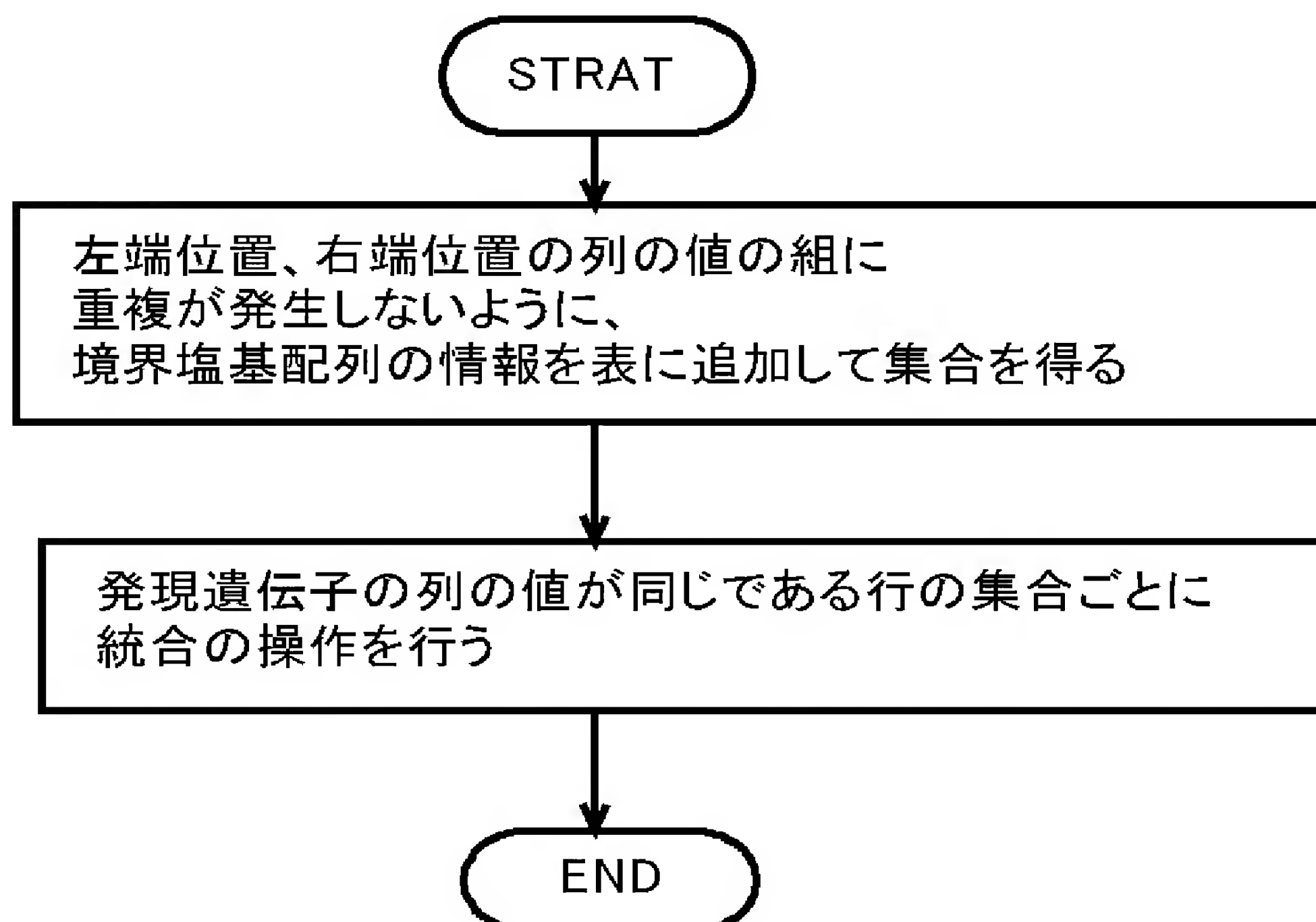
[図11]



[図12]

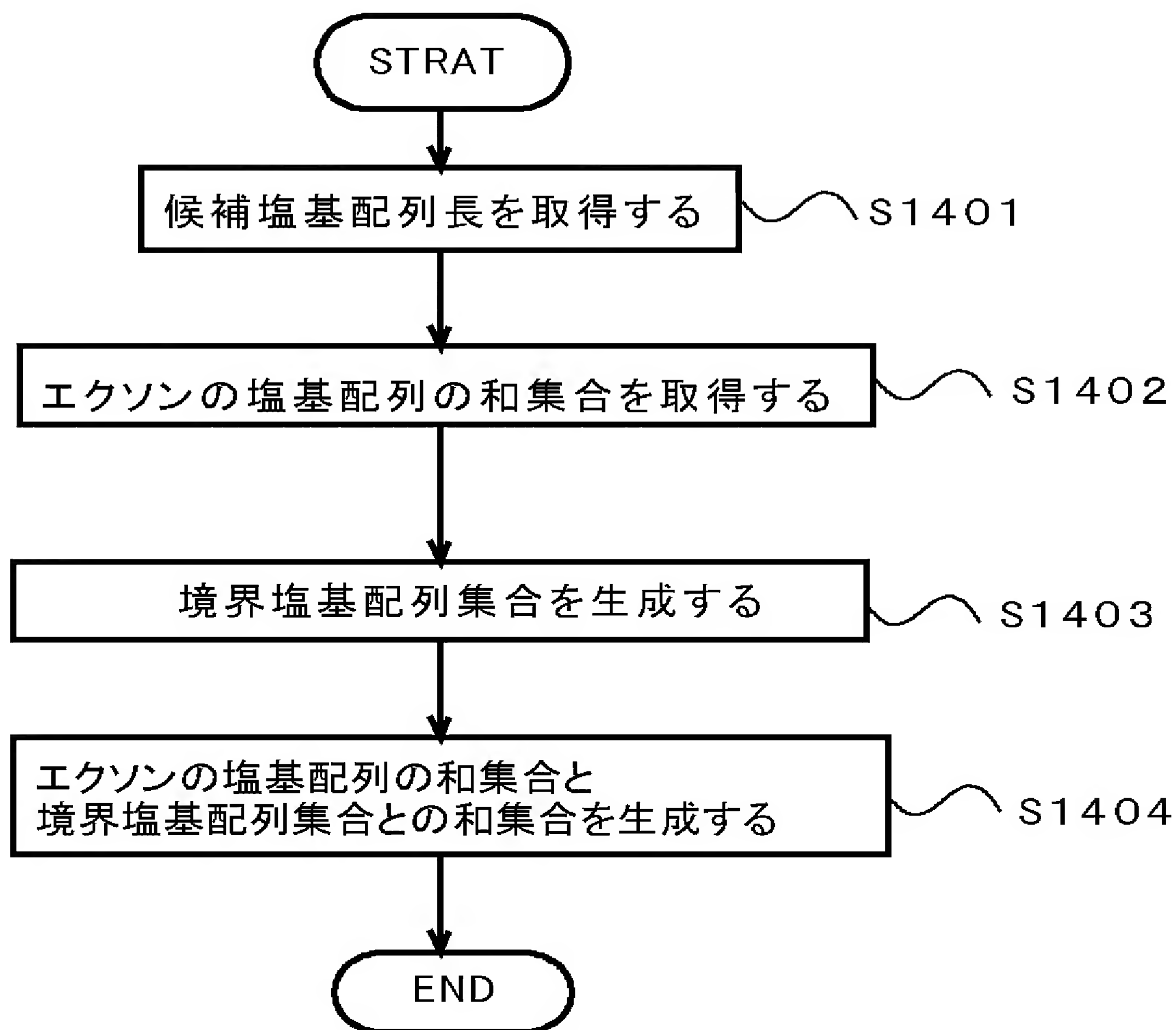
発現遺伝子	左端位置	右端位置
1101,1102,1103	2049	3659
1101,1102,1103	2048	3660
⋮	⋮	⋮
1101,1102,1103	2111	4520
1101,1102,1103	2111	4521
⋮	⋮	⋮

[図13]





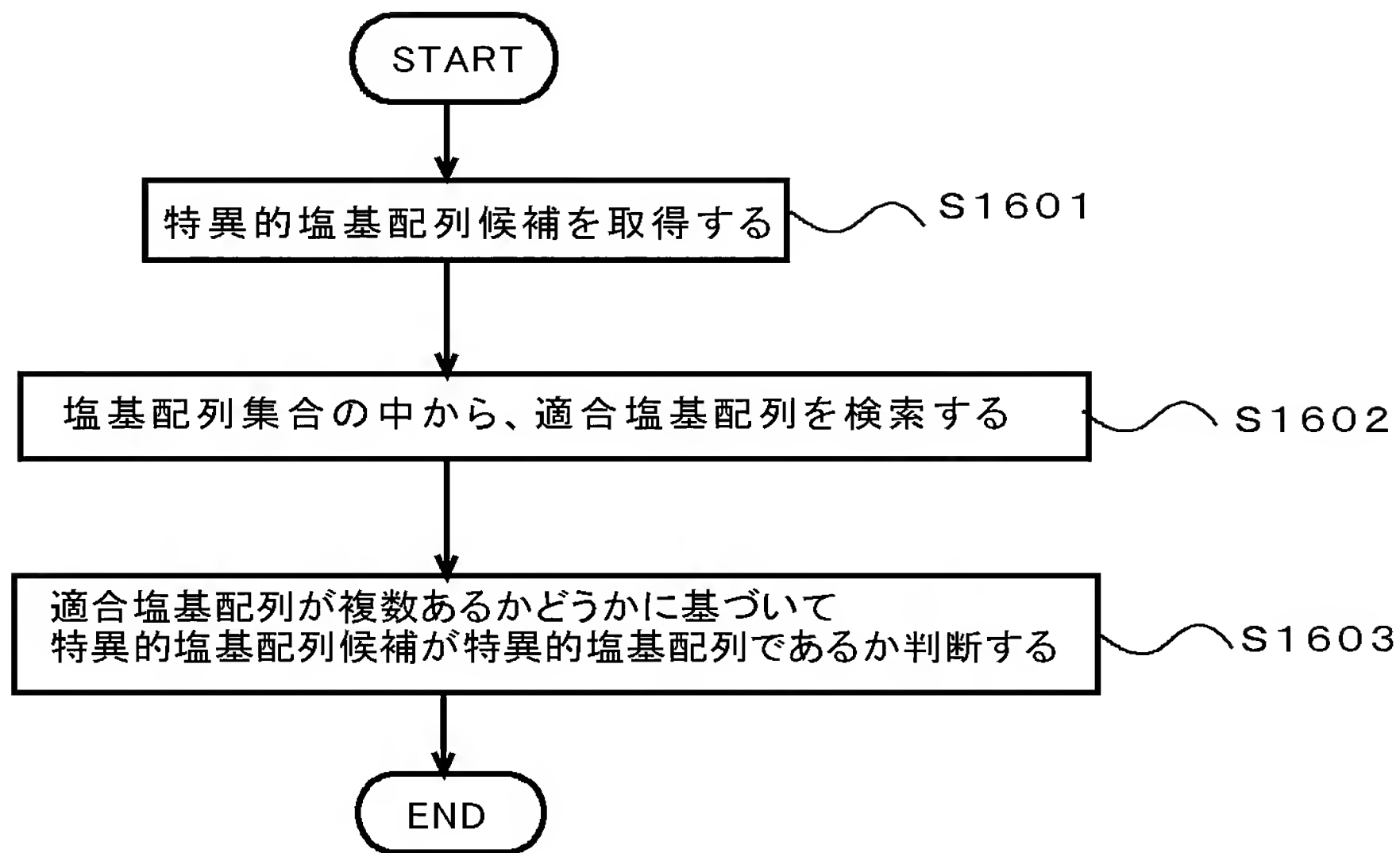
[図14]



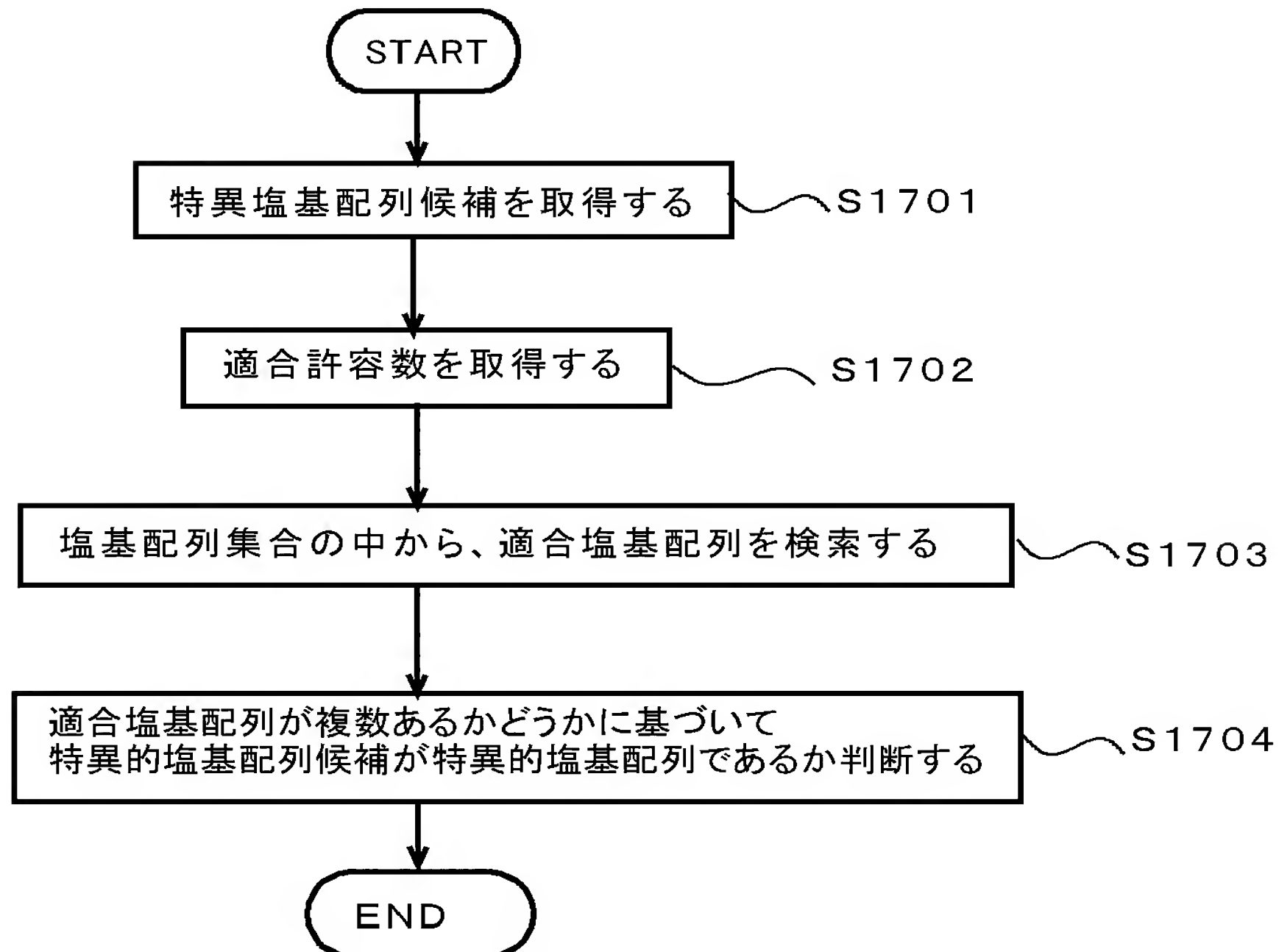
[図15]

左端位置	塩基配列
1024	ATCGTA
⋮	⋮

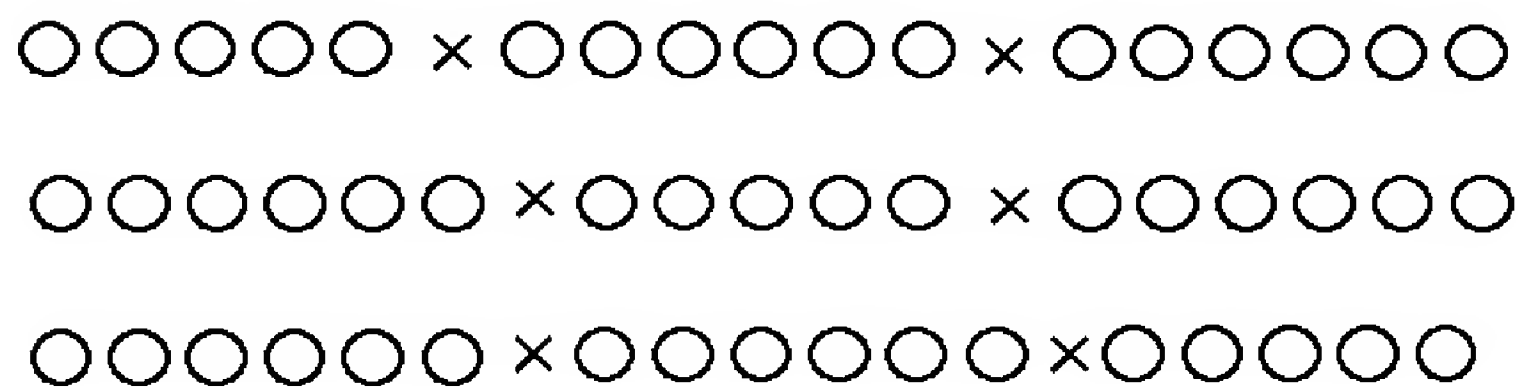
[図16]



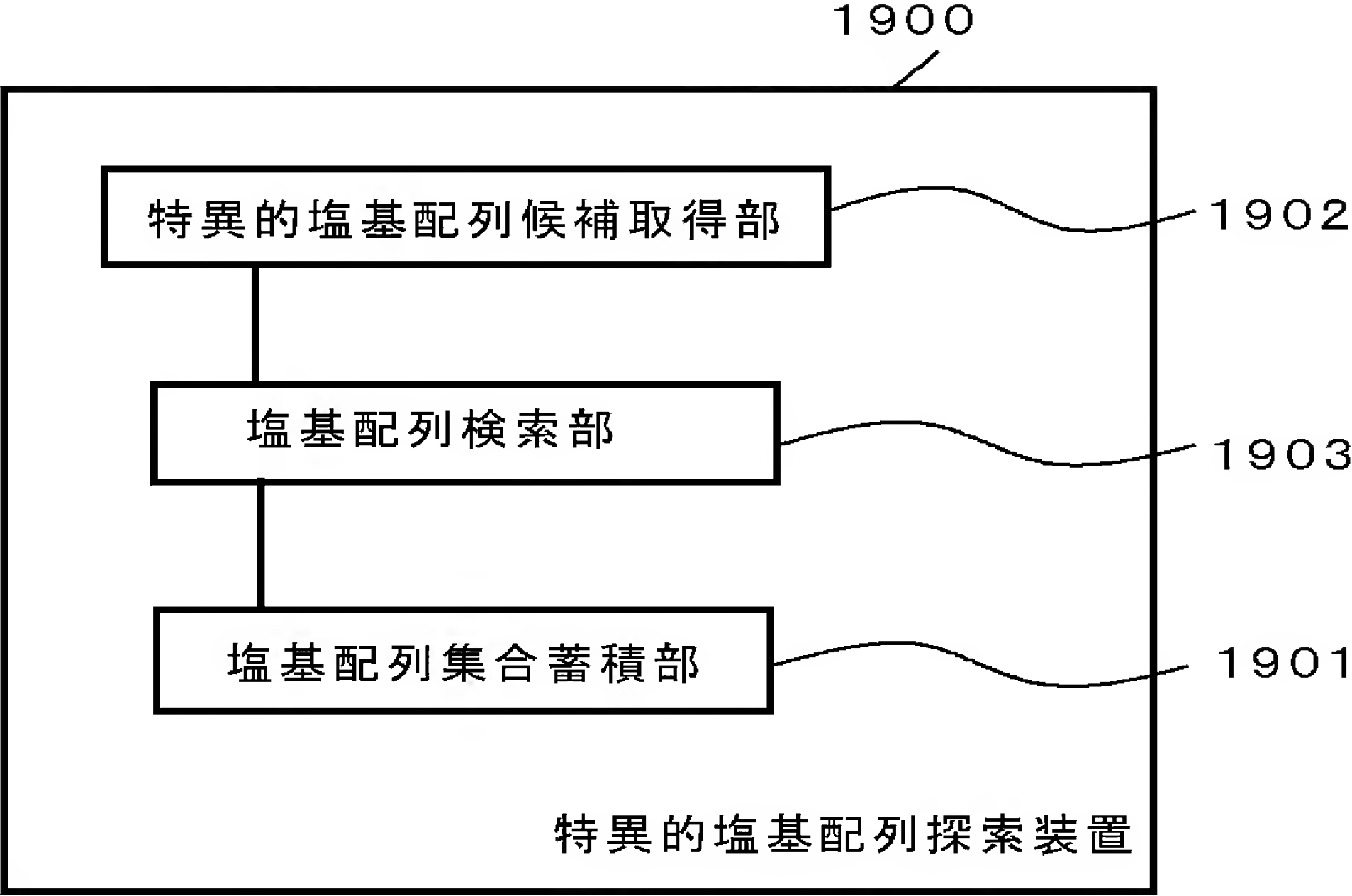
[図17]



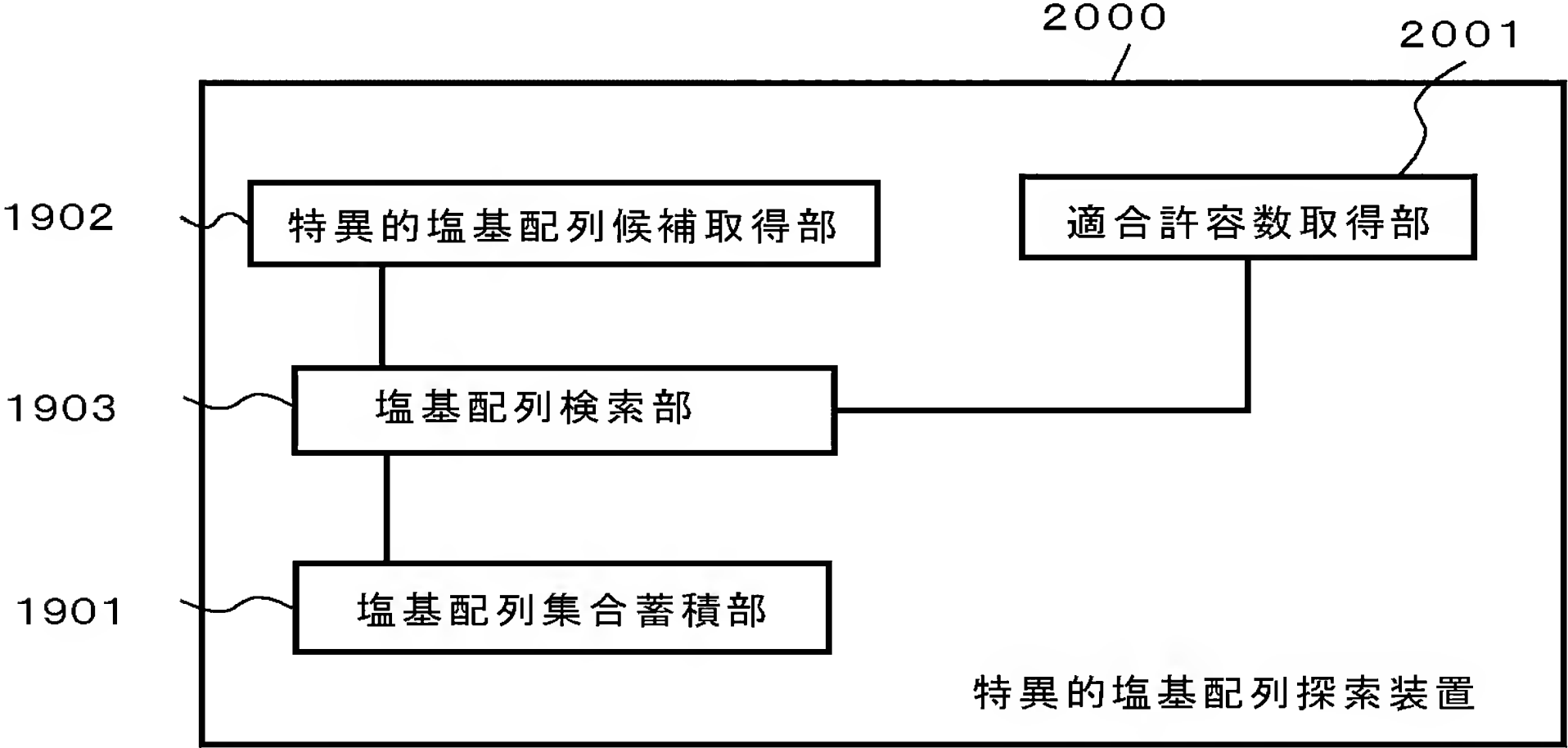
[図18]



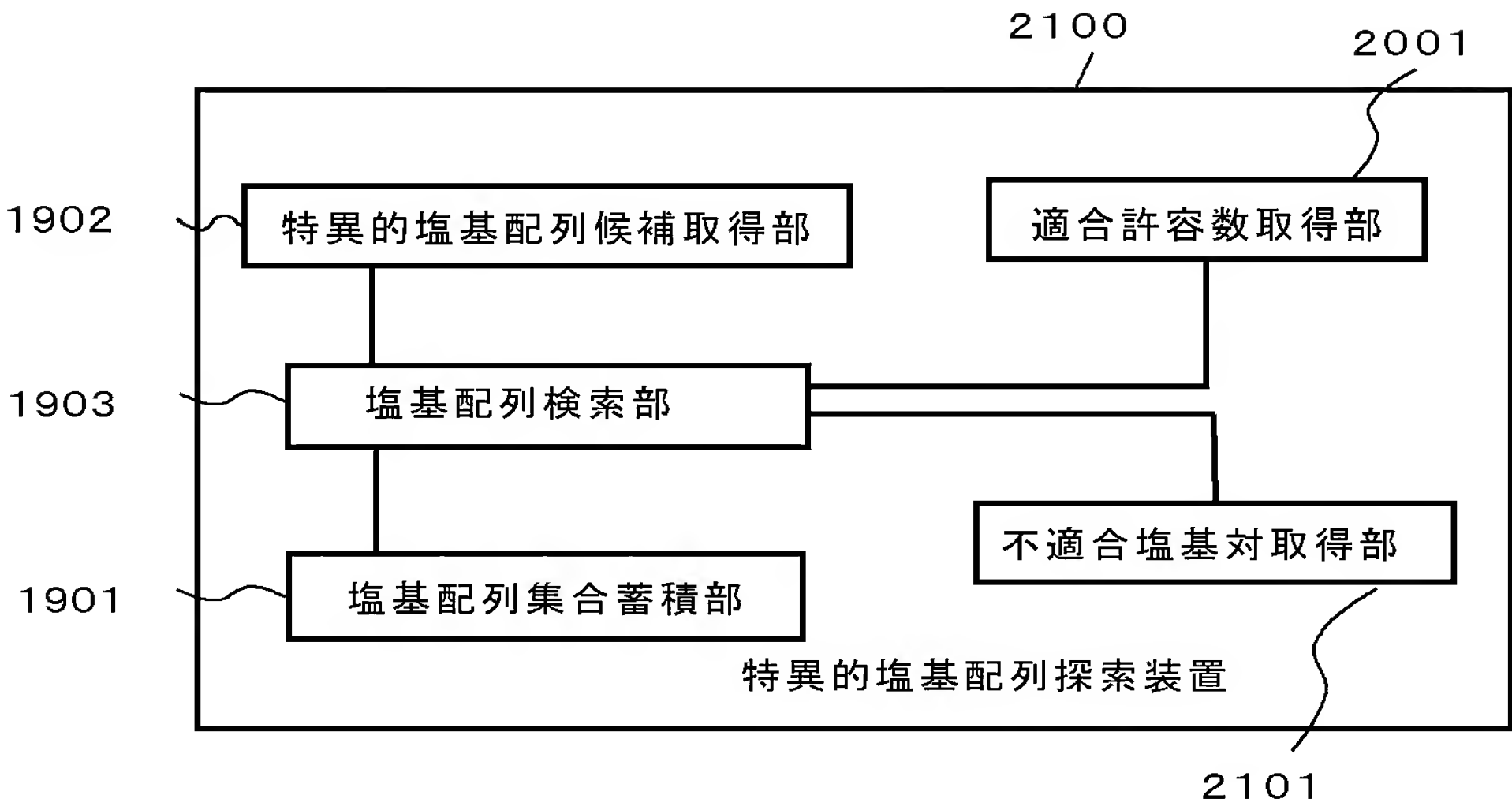
[図19]



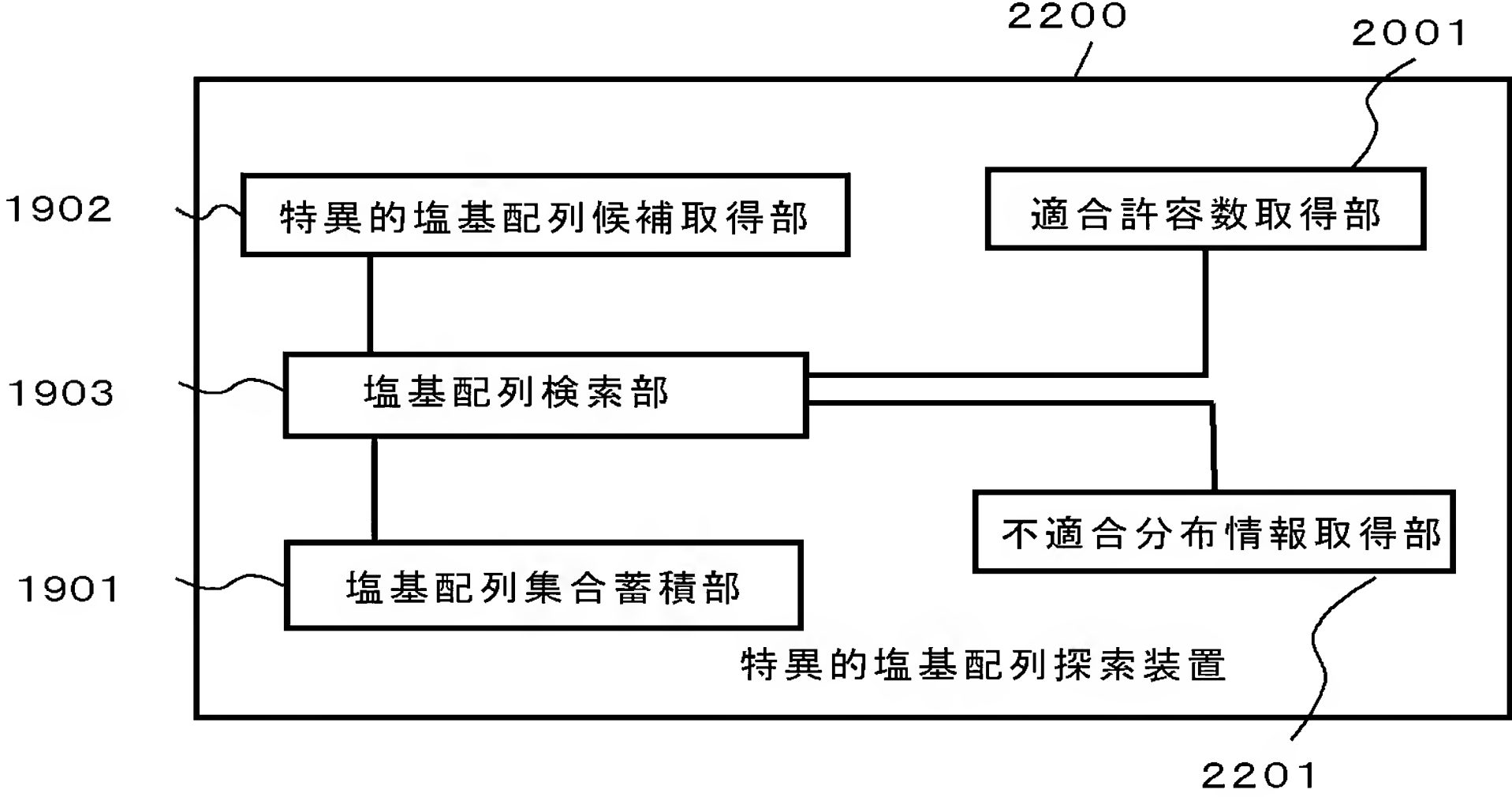
[図20]



[図21]



[図22]



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2005/005290

## A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl.<sup>7</sup> G06F19/00, C12M1/00, C12N15/09, C12Q1/68, G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl.<sup>7</sup> G06F19/00, C12M1/00, C12N15/09, C12Q1/68, G06F17/30

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2005

Kokai Jitsuyo Shinan Koho 1971-2005 Toroku Jitsuyo Shinan Koho 1994-2005

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PubMed, JSTPlus (JOIS)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	POSPISIL H. et al., EASED: Extended Alternatively Spliced EST Database, Nucl.Acids Res., Vol.32, Database issue, 01 January, 2004 (01.01.04), D70-D74	1-23
A	JP 2001-258568 A (Hitachi, Ltd.), 25 September, 2001 (25.09.01), Par. No. [0036]; Fig. 4 (Family: none)	1-23



Further documents are listed in the continuation of Box C.



See patent family annex.

\* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&amp;" document member of the same patent family

Date of the actual completion of the international search

27 May, 2005 (27.05.05)

Date of mailing of the international search report

14 June, 2005 (14.06.05)

Name and mailing address of the ISA/

Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.



A. 発明の属する分野の分類 (国際特許分類 (I P C)) Int.Cl. <sup>7</sup> G06F19/00, C12M1/00, C12N15/09, C12Q1/68, G06F17/30		
B. 調査を行った分野 調査を行った最小限資料 (国際特許分類 (I P C)) Int.Cl. <sup>7</sup> G06F19/00, C12M1/00, C12N15/09, C12Q1/68, G06F17/30		
最小限資料以外の資料で調査を行った分野に含まれるもの 日本国実用新案公報 1922-1996年 日本国公開実用新案公報 1971-2005年 日本国実用新案登録公報 1996-2005年 日本国登録実用新案公報 1994-2005年		
国際調査で使用した電子データベース (データベースの名称、調査に使用した用語) PubMed, JSTPlus(JOIS)		
C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	POSPISIL H, et al., EASED: Extended Alternatively Spliced EST Database, Nucl. Acids Res., Vol. 32, Database issue, 2004. 01. 01, D70-D74	1-23
A	JP 2001-258568 A (株式会社日立製作所) 2001. 09. 25, 段落【0036】, 図4 (ファミリーなし)	1-23
<input type="checkbox"/> C欄の続きにも文献が列举されている。 <input type="checkbox"/> パテントファミリーに関する別紙を参照。		
<p>* 引用文献のカテゴリー</p> <p>「A」特に関連のある文献ではなく、一般的技術水準を示すもの</p> <p>「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの</p> <p>「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)</p> <p>「O」口頭による開示、使用、展示等に言及する文献</p> <p>「P」国際出願日前で、かつ優先権の主張の基礎となる出願日の後に公表された文献</p> <p>「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの</p> <p>「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの</p> <p>「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの</p> <p>「&amp;」同一パテントファミリー文献</p>		
国際調査を完了した日 27. 05. 2005	国際調査報告の発送日 14. 6. 2005	
国際調査機関の名称及びあて先 日本国特許庁 (I S A/J P) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官 (権限のある職員) 岩間 直純 電話番号 03-3581-1101 内線 3562	5 L 9287